DTIC FILE COPY

(4)

Information and Computer Science

AD-A206 403

Exploring the Episodic Structure of Algebra Story Problem Solving *

Rogers Hall
Dennis Kibler
Etienne Wenger
Chris Truxaw

Irvine Computational Intelligence Project
Department of Information and Computer Science
University of California
Irvine, CA 92717

TECHNICAL REPORT

UNIVERSITY OF CALIFORNIA IRVINE

89 3 24 017

(4)

Exploring the Episodic Structure of Algebra Story Problem Solving *

Rogers Hall
Dennis Kibler
Etienne Wenger
Chris Truxaw

Irvine Computational Intelligence Project
Department of Information and Computer Science
University of California
Irvine, CA 92717

Technical Report 86–24

Revised: December 12, 1988

To appear in *Cognition & Instruction*

*Supported by Personnel Training and Research Division of the Office of Naval Research,contract N00014-85-K-0373.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Exploring the episodic structure of algebra story problem solving		5. TYPE OF REPORT & PERIOD COVERED Research Report 12/31/88 6. PERFORMING ORG. REPORT NUMBER UCI-ICS Tech Rep 86-24 (revised)
7. AUTHOR(s) Rogers Hall, Dennis Kibler, Etienne Wenger, Chris Truxaw		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0373
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Information & Computer Science University of California, Irvine, CA 92717		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Cognitive Science Program Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE January 1, 1989 13. NUMBER OF PAGES 90
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT *(of this Report)* Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*		
18. SUPPLEMENTARY NOTES To appear in *Cognition & Instruction*, 1989.		
19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)* mathematical problem solving model-based reasoning teaching representations algebra story problems analogical problem solving		
20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)* OVER		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT

This paper analyzes the quantitative and situational structure of algebra story problems, uses these materials to propose an interpretive framework for written problem-solving protocols, and then presents an exploratory study of the episodic structure of algebra story problem solving in a sizable group of mathematically competent subjects. Analyses of written protocols compare the strategic, tactical, and conceptual content of solution attempts, looking within these attempts at the interplay between problem comprehension and solution. Comprehension and solution of algebra story problems are complimentary activities, giving rise to a succession of problem solving episodes. While direct algebraic problem solving is sometimes effective, results suggest that the algebraic formalism may be of little help in comprehending the quantitative constraints posed in a problem text. Instead, competent problem solvers often reason within the situational context presented by a story problem, using various forms of "model-based reasoning" to identify, pursue, and verify quantitative constraints required for solution. The paper concludes by discussing the implications of these findings for acquiring mathematical concepts (e.g., related linear functions) and for supporting their acquisition through instruction.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Exploring the Episodic Structure of Algebra Story Problem Solving

Rogers Hall, Dennis Kibler,
Etienne Wenger and Chris Truxaw

Department of Information and Computer Science
University of California at Irvine

Send proofs to:
Rogers Hall
Department of Information and Computer Science
University of California, Irvine, CA 92717

ABSTRACT

This paper analyzes the quantitative and situational structure of algebra story problems, *uses these materials to propose an interpretive framework for written problem solving protocols*, and then presents an exploratory study of the episodic structure of algebra story problem solving in a *sizable group of mathematically competent subjects*. Analyses of written protocols compare the strategic, tactical, and conceptual content of solution attempts, looking within these attempts at the interplay between problem comprehension and solution. Comprehension and solution of algebra story problems are complimentary activities, giving rise to a succession of problem solving episodes. While direct algebraic problem solving is sometimes effective, results suggest that the algebraic formalism may be of little help in comprehending the quantitative constraints posed in a problem text. Instead, competent problem solvers often reason within the situational context presented by a story problem, using various forms of "model based reasoning" to identify, pursue, and verify quantitative constraints required for solution. The paper concludes by discussing the implications of these findings for acquiring mathematical concepts (e.g., related linear functions) and for supporting their acquisition through instruction.

Confronted with an algebra story problem, a student faces a fundamental sort of "ill structured problem" (Newell, 1969; Simon, 1973). The problem text gives information about initial and goal states, but state transition operators taking the text into a quantitative solution are hardly well defined. Even assuming the student has an adequate grasp of mathematical principles and operators within the formalisms of arithmetic and algebra (e.g., the distributive property of multiplication over addition or using algebraic substitution), a solution to the presented problem is often obvious only in retrospect. Rather than searching for a solution path in a well-defined space of representational states, the problem solver is more likely to be searching among a space of alternative representations in an attempt to make the problem routine or familiar. Omitted or incorrectly introduced constraints within the problem representation can lead to prolonged and often meaningless calculations, and may encourage otherwise sophisticated problem solvers to give up entirely. Information-processing models of ill-structured problem solving remain elusive.

This state of affairs might be puzzling but acceptable if algebra story problems were transient disturbances in the secondary school curriculum. However, these problems recur as a general task throughout the mathematics curriculum and are even found in the quantitative sections of entrance examinations for professional schools. If prevalence alone is an insufficient basis for study, the unique role of these problems in bringing mathematical formalisms into contact with everyday experience recommends them highly. Viewed from within the classroom, story or "applied" problems provide students with an opportunity to validate acquired mathematical abstractions in more familiar domains (e.g., traveling or shopping). Viewed in a wider context, these problems may provide a curricular microcosm of a central pedagogical problem: transfer of training from the algebra classroom to students' later educational or life experiences. Interpretations derived from either vantage are controversial. For example, we have anecdotal evidence that these problems are avoided by some teachers as being too difficult for both students and teachers. On the other hand,

studies of mathematics in practice suggest that "real world" curricular materials may have little correspondence with mathematical problems or their solution in "real life" (Lave, 1986, 1988). For psychologists and educationalists alike, the problem is to determine how applied problems are solved by competent problem solvers and how acquisition of that competence might be supported.

[Insert Table 1 about here.]

Algebra story problems of the sort shown in Table 1 have been studied extensively by cognitive and educational psychologists, both as a representative task for mathematical problem solving (e.g., Hinsley, Hayes, & Simon, 1977; Kilpatrick, 1967; Mayer, Larkin, & Kadane, 1984; and Paige & Simon, 1968) and as experimental materials for studies of transfer (e.g., Dellarosa, 1985; Reed, 1987; Reed, Dempster, & Ettinger, 1985; and Silver, 1979, 1981). Many studies treat problem solving as an opaque process with an inspectable output (i.e., correct or incorrect) and duration. Manipulations in problem content or presentation are introduced, performance data are collected, and inferences are drawn concerning hypothetical problem-solving mechanisms. In contrast, much as in Kilpatrick's early work (1967) and subsequent studies of mathematical problem solving by Lucas (1979) and Schoenfeld (1985), we have chosen instead to present subjects with representative problems and then to observe and analyze their uninterrupted responses in some detail. This approach trades experimental control over the problem solving setting for a richer (albeit interpretive) view of problem-solving activities. In addition to finding whether or not a subject has gotten a problem "right," we are at least partially able to explore the solution strategies that subjects select and their tactical course in achieving solutions, right or wrong. We find this a useful approach to characterizing what competent problem solvers actually do when solving these problems (i.e., a succession of strategic and tactical efforts). This characterization is a necessary first step towards finding methods for supporting acquisition of competent problem solving behaviors.

When describing models of algebra story problem solving, we will distinguish between the *generative* and *predictive* capacity[1] of models (computational or otherwise) as successive approximations to a robust instructional theory. A model with generative capacity uses an expressive language for describing problems and their solutions to produce descriptions of problem solving activity that obey certain constraints. For example, given a language that is adequate for expressing arbitrary algebraic expressions, we might like to generate only those expressions that reflect mathematical relations stated directly in a story problem text. For various instructional purposes, this is an improvement over generating *all* syntactically permissible algebraic expressions, but it falls well short of addressing typical instructional problems — e.g., why or how has a student generated some particular algebraic expressions? This sort of predictive capacity will require considerable extensions to the expressive language (e.g., a notation for intermediate representational states) and to constraints that restrict the process for generating algebraic expressions (e.g., a vocabulary of justifications for a subject's choices among alternative problem solving activities). Given a sufficiently expressive language and an appropriate set of constraints, a model may generate descriptions that correspond closely with students' activities. When this correspondence is of high fidelity — i.e., the model answers questions of why or how in a psychologically plausible fashion — it can be used to support a variety of important instructional tasks. For example, a predictive model of algebra story problem solving might be used to diagnose students' errors, to plan tutorial activities, or even to provide basic instructional materials.

Work reported in this paper approaches a predictive model by presenting descriptive languages for problem-solving activities, examining constraints that arise from interactions between these languages, and then exploring problem solving behaviors observed in a sizable group of competent problem solvers. In the first section of the paper we examine some

[1] We are not arguing for explanatory adequacy in the sense usually reserved for linguistic theories (Chomsky, 1965). The models discussed in this paper approach descriptive adequacy but do not yet propose stronger constraints on acquiring problem solving competence.

basic materials out of which algebra story problems and their solutions can be constructed. Our working hypothesis is that in order to generate a solution-enabling representation of a problem, reasoners must assemble quantitative constraints under the guidance of their understanding of the *situational context* presented by the story problem. This context serves not only as a vehicle for the quantitative problem, but also as a framework for justifying the *existence of quantitative constraints and their interrelationships. Accordingly, we examine* the quantitative and situational structure of typical algebra story problems, and then use representative problems in the exploratory study.

In later sections of the paper we analyze the written protocols of 85 upper division computer science undergraduates who were instructed to show their work when solving four, representative algebra story problems. An interpretive framework is developed in which a written solution attempt is divided into a series of coherent problem solving episodes. Each of these episodes is coded along a set of categories reflecting strategic and tactical role, conceptual content, manipulative or conceptual errors, and relationship to surrounding episodes. *Exploratory analyses of the scored protocols provide evidence for the frequency* with which various problem solving behaviors occur within subjects' solution attempts, the content and outcome of the "final episodes" during which subjects conclude their efforts, and the role that "model based reasoning" plays in solution attempts. One of our central findings is that competent problem solvers frequently engage in problem solving activities "outside" of the traditional algebraic formalism. These activities, based on an analysis of protocol results, often take the form of constructive and elaborative problem solving inferences within the situational context presented by an algebra story problem. These findings are interpreted as evidence for a model of quantitative problem solving in which *mathematical formalisms (e.g., algebraic expressions) provide a sometimes useful tool for* comprehending quantitative constraints. In the discussion section, we relate this model to existing accounts of mathematical problem solving, and then consider the implications of

these findings for acquiring mathematical concepts (e.g., related linear functions) and for supporting their acquisition through instruction.

PROBLEM STRUCTURE

Before presenting our exploratory study, we examine the domain of algebra story problems at two levels of abstraction: the ***quantitative structure*** of related mathematical entities and the ***situational structure*** of related physical entities within a problem. The central activity in our model of problem solving is to find convergent constraints through constructive elaboration of a problem representation. Structural abstractions examined in this section give two basic materials for such a constructive process. Ultimately, these and other levels of analysis may provide a relatively complete domain "ontology" (Greeno, 1983) for algebra story problems and other situations that give rise to mathematical problem solving. For the purposes of this paper, we want to identify materials that can provide a descriptive vocabulary for behavioral observations presented *in later sections* and can assist our intuitions in framing a model of problem solving, learning, and teaching within this domain. These materials can play several roles: as a description of the task of solving algebra story problems, as a hypothetical account of the representations held by subjects during the solution process, and as an illustrative medium for teaching. This section focuses on task and representational issues; the utility of quantitative and situational structures in education is examined in the discussion section.

Quantitative structure

By the quantitative structure of algebra story problems, we mean the mathematical entities and relationships presented or implied in the problem text. A particular problem has a "structure" at this level of analysis to the extent that the relationships between mathematical entities take a distinguishable form when compared with other algebra story problems. Of course, there might be many ways of characterizing the quantitative structure of an

arbitrary problem or group of ostensibly related problems — e.g., as algebraic equations or as matrices of coefficients. Bobrow (1968) uses algebraic equations as a canonical internal representation of meaning for story problem texts, while Reed et al. (1985) use equations to define the *a priori* similarity of problems and their solution procedures. The language of algebraic equations may be sufficient for analyzing the task of algebraic manipulation, but it is less useful when the analysis is to include what students actually understand and use while learning to solve algebra story problems.

A network language of quantitative entities. We start with a conceptual framework originally proposed by Quintero (1981; Quintero & Schwartz, 1981) and later extended by Shalin & Bee (1985) and Greeno (1985, 1987; Greeno, Brown, Foss, Shalin, Bee, Lewis, & Vitolo, 1986). The framework serves all three roles mentioned above: as an analysis of task structure, as a hypothetical account of subjects' representations of algebra problems, and as an instructional medium. Our interest in this work is twofold. First, we will use the framework as a means for describing constraints essential for problem solution, although several additions to the framework would be necessary *for it to serve* as a representational hypothesis. Second, we will employ some aspects of the framework to describe how an arbitrary pair of problems might be considered similar for problem solving purposes.

Shalin & Bee (1985) describe the mathematical structure of an algebra story problem as a network consisting of quantitative elements, relations over those elements, and compositions of these relations. Quantitative elements are divided into four basic types: an **extensive** element denotes a primary quantity (e.g., some number of miles or hours), an **intensive** element denotes a map between two extensives (e.g., a motion rate relates time and distance); a **difference** element poses an additive contrast of two extensives (e.g., one time interval is 2 hours longer than another); and a **factor** element gives a multiplicative comparison of two extensives (e.g., one distance is twice another). Composing these elements according to their type yields quantitative relations. A quantitative relation is

defined as an arithmetic operation (*i.e.*, *addition*, *subtraction*, *multiplication*, or *division*) relating three quantities. For example, the fact that a train traveling 100 km/h for 5.5 hours covers a distance of 550 km can be expressed as a relational triad over two extensives (550 kilometers and 5.5 hours) and a single intensive (100 km/h) as shown in Figure 1. Each element is presented graphically as a box containing several expressions. The shape at the top of the box designates element type — e.g., a rectangular top designates an extensive, a triangular top an intensive.

[Insert Figure 1 about here.]

As an additional level of structure, relational triads can be composed by sharing various quantities to yield "problem structures." These are quantitative networks describing typed quantities and *constraints* among them. As shown with solid lines[2] in Figure 2(a), a single quantitative network can be used to graphically represent the problem of trains traveling in opposite directions (problem *MOD* from Table 1). Sharing a common time, two rates combine through multiplicative triads to yield parts of the total distance. These parts are combined in an additive triad to give a single extensive quantity representing the total distance. Figure 2(b) shows a quantitative network corresponding to the round trip (*MRT*) problem. In both networks, the term "output" serves as a generalization over distance and work.

Taken together, quantities, relations and structures provide a language for describing the quantitative form of particular algebra story problems. While a variety of equivalent graphical languages might be used (e.g., parse trees for arithmetic expressions), this language gives explicit representational status to mathematical entities, associates a quantitative type with each, and incorporates a metaphorical sense of storage for holding semantic information (e.g., textual phrases) and intermediate calculations. Constraints on the arithmetic composition of typed quantitative entities restrict the space of possible quantitative

[2]Portions of the network in dashed lines will be discussed shortly.

relations (Greeno *et al.*, 1986). For example, the multiplicative composition of intensive and extensive quantities (rate and time) in Figure 1 is allowed, while an additive composition of the same quantities would be disallowed. Greeno (1987) points out that constraints are also available from compositional restrictions on the units of measurement for quantities[3], although the network language does not presently incorporate these constraints. Finally, the interconnectivity of a quantitative network supports a form of written algebraic calculus. Expressions can be propagated through the network with the goal of finding convergent constraints on the given unknown.

[Insert Figure 2 about here.]

Quantitative networks provide a visually accessible notation for comparing the structure of different algebra story problems. However, the notation and compositional constraints do not specify which of the permissible quantitative structures a subject might generate when solving an algebra story problem. For example, the quantitative network shown with solid lines in Figure 2(a) describes the opposite direction problem *after* several crucial inferences have occurred: component distances have been inferred within the total of 880 kilometers, and a single extensive quantity for travel time has been correctly inserted in the network. For the same problem, consider elaborating the quantitative network to include network components shown with dashed lines in Figure 2(a). We might imagine a subject inferring that the given rates can be added. The resulting combined rate (160 km/h), when multiplied by the unknown time, gives the total distance without adding constituent distances. During empirical studies with this and similar problems, we find considerable variety in the solution approaches taken by different subjects as well as by individual subjects within a single problem solving effort.

The networks shown in Figure 2 are idealized graphical representations of problem

[3]An instructional tool developed by Schwartz (1982) enforces unit constraints to help users avoid irrelevant calculations, particularly when using intensive quantities. Thompson (1988) combines quantitative networks and unit constraints in another tool named "Word Problem Assistant."

structure as they might be constructed by problem solvers who understand the quantitative network language and are able to use the language to comprehend and solve algebra story problems. These networks give a particular quantitative representation, but their content is largely the result of inferential processes that draw on other knowledge sources. These processes may include: recognizing quantitative entities directly contained in or implied by the problem text, composing these entities into local relational structures, composing relational substructures into larger problem structures, recognizing familiar substructural arrangements, and detecting when constraints are sufficient for solution. The results of each action lie within the quantitative formalism for which Shalin & Bee's (1985) framework provides a functional description. However, the enablement conditions for these actions or the knowledge sources that support them lie partly outside the formalism. These issues are explored further when we consider the situational structure of problems.

Quantitative networks as problem classes. Quantitative networks provide an analytic tool for examining aspects of quantitative similarity. At the level of entire problems, *this approach gives a stronger basis for mathematical similarity than simply noting common* equations. At a more fine grained level, there may be significant areas of substructural isomorphism in globally dissimilar problems.

The problems from Table 1 can be grouped into structurally similar pairs as follows: *MOD/WT* and *MRT/WC*. Each problem in a pair is a "quantitative isomorph" of the other, as shown graphically in Figure 2. In the *MOD/WT* pair, extensives for kilometers traveled correspond with those for parts of a job completed (output1 and output2). In the *MRT/WC* pair, a round trip travel extensive corresponds with an extensive for boxes filled and then checked (output). Comparing problems within each pair, extensive and intensive quantities play identical roles in the surrounding network structure. However, when comparing problems across pairs, structural roles of similar quantitative entities change or are even reversed. For example, the additive extensive relation for combined distance (or work

output) in Figure 2(a) is locally similar to the additive extensive relation for combined *time in Figure 2(b), but these relations play very different roles in their overall quantitative* structures. In general, a specific quantitative network defines an equivalence class of algebraic problems, each of which may have a different situational instantiation. Of course, being directly similar in form does not mean that problems must be solved in the same way. Figure 2 presents the quantitative structure of problem materials required for a quantitative solution. We could as well depict the quantitative structure of intermediate representational states in subjects' solutions, an exercise that sometimes leads to a surprising sequence of graphical images as various conceptual errors are introduced or repaired.

Turning to a finer grained level of comparison, we can identify classes of problems that are similar to each other by sharing particular quantitative substructures. A substructure is a subgraph within a larger quantitative network consisting of stated quantities, inferred quantities, and relationships among these quantities. For example, "current" problems are similar at a quantitative level because they contain an additive relationship between the rate of the vehicle (steamer, canoe, etc.) and the rate of the medium in which it travels (current, tide, etc.). While other aspects of the quantitative structure for a pair of current problems can be dissimilar, such a shared substructure may contribute to subjects' estimates of problem similarity. As in the results of Hinsley *et al.* (1979), similarity judgments at the level of "river" problems may appear an educational failure: problem solvers acquire content specific categorizations when the true pedagogical goal is to facilitate their learning of mathematical forms. Another interpretation is that quantitative substructures are learned through instruction and problem solving experience and thus form part of the underlying competence in this domain. Since particular substructures are correlated with problem types, the resulting categorizations appear overly content specific. However, there may be a functional or pragmatic basis for learning these problem classes: despite dissimilarity of overall mathematical structure, shared quantitative substructures require similar

partial solution strategies.

Situational structure

The quantitative network formalism does not attempt to account for the problem structures that subjects actually generate during problem solving, although some constraints are placed on combining quantitative types into relational triads. In this section, we examine another level of abstraction — the *situational structure of a story problem* — as a source of additional constraint when subjects construct a solution enabling representation of an algebra story problem. Our view of the situational structure of an algebra story problem is not synonymous with what other researchers have called "surface content." Although surface materials like trains, buses, or letters are important problem constituents, and may be particularly so for novice problem solvers, we will not focus on these materials.

Instead, we present a language for describing the situational structure of "compound" algebra story problems involving related linear functions, and use the language in a detailed examination of problems involving motion or work[4] (see Table 1). As with the quantitative network formalism, our language for describing the situational structure of problems can play several roles: as an analysis of problem structure, as a hypothetical cognitive representation, or as an educational medium. *Here we develop a relational language for describing* problems, argue for its utility in generating key problem-solving inferences, and then use the language to present a viewpoint on the space of possible algebra story problems that is complementary to problem classes based on quantitative structure. In later sections of the paper, we also use the language to help interpret various activities observed in an exploratory study of algebra story problem solving and then to consider the educational implications of these findings.

A relational language of situational contexts. We present the basic terms of our

[4] Motion and work are frequently used as the setting for story problems in algebra texts, comprising 20% of an extensive sampling by Mayer (1981).

relational language first, followed by an example of their use shown in Figure 3. Compound motion and work problems are assembled around related events — e.g., traveling in opposite directions, working together, riding a bus and walking, or filling envelopes and checking them. In each event, an agent engages in activity that produces some output (distance or work) over a period of time. Hence, output and time are the basic **dimensions** that organize story activities. These activities start and stop with particular times, locations, or work products that *can be modeled as* **places** *along the appropriate dimension*. Places that bound an activity define particular **segments** of output or time, and these segments can be placed in relation to each other within a common dimension[5]. Rates of motion or work give a systematic correspondence between dimensions of output and time, and using rates in the solution of a quantitative problem requires a strategy for integrating these dimensions. Arranging output and time dimensions orthogonally gives a rectilinear framework in which rate is a two-dimensional entity. We model these rate entities as **inclines** that associate particular output and time segments. Relational descriptions involving typed dimensions, places, segments, and inclines provide a language for expressing the **situational context** of an individual problem.

[Insert Figure 3 about here.]

The situational context of problem *MOD* (from Table 1) is shown in Figure 3. Parts (a) and (b) of the figure show place and segment representations for output (distance) and time, while part (c) of the figure shows an orthogonal integration of these dimensions with time on the vertical axis. In part (a) of the figure, trains traveling in opposite directions from the same station provide two spatial segments (Distance 1 and Distance 2) sharing a place of origin (S) but with unknown places for destinations. These segments are *collinear* and oriented in *opposite* directions. Since trains leave from the same place of origin, these distance segments are also *adjacent* and can be arranged within the horizontal dimension

[5] Segment relations within a dimension are similar to Allen's (1983, 1984) relational descriptions of temporal intervals.

shown in part (c) of the figure. In part (b) of the figure, trains leave at the same time (t0) and are separated by *880 kilometers at some later time, providing time segments* (Time 1 and Time 2) that are *congruent* (i.e., coinciding at all points) when arranged within the vertical dimension. We assume collinearity and the same directional orientation for all time segments. Representing rates of travel as two-dimensional inclines, part (*c*) of the figure puts particular instances of output and time in correspondence (e.g., 60 versus 100 kilometers in the first hour of travel). Inclines can either represent a concrete situation, as shown here, or an invariant relation between output and time dimensions. Treating rate as an *invariant* relation approaches the algebraic sense of rate as a linear function. Each interpretation enables different problem-solving activities, discussed below.

Problem-solving inferences based on situational contexts. Before using this relational language to describe a space of situational contexts for algebra story problems, we briefly consider its utility as a representation for problem solving. First, we describe how a representation of situational context like that shown in Figure 3 could be constructed; second, we consider how this relational description might be useful for problem comprehension and solution. Both are ongoing research questions that touch on the role of our situational language as a representational hypothesis and an instructional medium.

On the issue of how these representations might be constructed (either spontaneously or as an educational exercise), we propose a series of *constructive inferences* that operate on a case frame representation[6] of the events described in the text of a compound algebra story problem. *These inferences build a situational model of the problem by assembling* a relational description of a particular situational context. Assuming the case frames contain roles that specify typed places and segments (e.g., the starting location versus the starting time), we can model these roles as situational places and segments within output and time dimensions. From these initial situational entities, a series of elaborative inferences

[6]See Brachman (1979) for a review of related representation schemes and Kintsch & Greeno (1985) for an example of a case frame representation for the text of word arithmetic problems

identify places and segments *implicit in the problem statement* and relations over segments within each dimension. What results is a relational description of situational context as in Figure 3. Constructive inferences that assemble a relational description of situational context are similar to the comprehension strategies that Kintsch & Greeno (1985) use to take propositional encodings of arithmetic word problems into a set based representation.

On the issue of utility, we suspect that segment relations within situational dimensions *support the construction of quantitative representations like the networks of Shalin* & Bee (1985). For example, knowing that spatial segments are *collinear* and *adjacent* while times are *congruent* supports two useful problem-solving inferences in problem *MOD*: constituent distances can be added to yield a total distance, and the rates of each train can be added to give a combined rate. The first inference is a necessary quantitative constraint for solution, while the second inference effectively compresses the compound problem into a simpler problem which can be solved without extended algebraic manipulation. These are precisely the inferences about problem structure that were not accounted for in our examination of quantitative structure. For example, the network components shown with dashed lines in Figure 2(a) would result if a student decided to add motion or working rates. Hence, in addition to constructive inferences that build a situational context, there are also *constraint-generating inferences* that take descriptions of situational structure into quantitative relations. Each inference about a quantitative constraint, supported by relevant *situational relations*, gives a substructural component in a larger set of constraints that may enable a solution.

It is also possible to use dimensions, places, segments, and inclines directly in a solution attempt by treating these *representational entities* as a model *of the problem situation*. We will develop a general account of *model based reasoning* as a problem solving tactic here. Following sections introduce operational categories for interpreting this tactic within the structure of written protocols and give an empirical account of its use and *consequence in*

algebra story problem solving.

[Insert Figure 4 about here.]

Placed within a *single dimension to model time or output*, segments provide an explicit spatial representation that enables a variety of problem solving operations like "copying," "stacking," "comparing," or "decomposing" their one-dimensional extent. Similarly, using inclines as models of rate enables operations like "joining" or "scaling" their two dimensional extent. *Joining*, shown in part (a) of Figure 4, places copies of the concrete incline along the diagonal in an iterative fashion. Scaling, in part (b) of the figure, treats the incline as an invariant relation by estimating the extent of a segment in one dimension and then projecting that value through the incline to generate an associated extent in the other dimension. Each operation is based on a different interpretation of rate as a relation across dimensions, and each coordinates operations on associated segments within single dimensions.

Both join and scale operations enable problem solving by model-based reasoning without requiring algebraic representation. Figure 5 shows solution attempts using join and scale operations on the opposite direction motion problem (MOD). Treating *inclines* as concrete entities in part (a) of the figure, the join operator enables an iterative simulation over five successive one hour increments in the time dimension. These correspond to intermediate states in a two-dimensional model of the problem, successively constructed and tested against the given constraint of being 880 kilometers apart after a common interval of time. Treating inclines as invariant relations in part (b) of the figure, the scale operator enables a heuristic estimate of the problem's final state by choosing five hours as the time at which the trains will be 880 kilometers apart and projecting this choice of a common time through each incline to find associated distance segments. In both solution attempts, spatial relations within the two dimensional model support and organize relatively simple quantitative operations like addition, multiplication, and value comparison. Thus, even without

utilizing the metric qualities that such a model might afford (e.g., testing whether adjacent distance segments precisely "fill" the composed 880 kilometer segment), model based reasoning can lead to a solution without explicitly constructing an algebraic *representation* of the problem.

[Insert Figure 5 about here.]

While entities and operations in model based reasoning can support solution attempts directly, they also provide a vocabulary of problem solving activities that could be used to construct an algebraic representation. For example, introducing a variable, t, as a label on the unknown common time in part (b) of Figure 5, we can use the scale operator to project that variable into expressions for labels on each distance segment in the horizontal dimension. Since these segments are *adjacent* and must fill the given combined distance of 880 kilometers, addition of label expressions in the horizontal dimension gives an algebraic expression for the combined distance, $100t + 60t = 880$. Thus, model based reasoning operations can also participate in *constraint generating inferences* described earlier.

In general, inferences in model based reasoning correspond to relatively opaque operations in the algebraic formalism (e.g., distribution of a product). *Their spatial character* and granularity may provide an accessible problem solving medium for subjects who are newcomers to the algebraic formalism. In addition, the results of these operations could justify more abstract activities in an algebraic or quantitative network representation, allowing problem solvers to verify quantitative constraints or results about which they are uncertain. Evidence for these hypothetical roles of model based reasoning, even in competent problem solvers, is presented in the sections that follow.

Situational contexts as problem classes. Beyond their role as a representational hypothesis or an instructional medium, situational contexts provide a viewpoint on the space of possible compound algebra story problems that is complementary to the problem classes provided by quantitative structure. Even if we restrict analysis to compound mo

tion problems in which movement must be *collinear* and directed, a variety of situational contexts are possible. Taking two *collinear* distance segments we can select from a set of spatial relationships (e.g., *congruent* or *adjacent*) and combine this selection with directional orientation (e.g., *same* or *opposite*) to yield a distinct spatial situation. Also selecting a relation between time segments (e.g., *congruent* or *adjacent*), we can combine segment relations for distance and time dimensions to yield a particular situational context for a compound motion problem. For example, problem *MOD* has *adjacent* distance segments oriented in *opposite* directions and has *congruent* time segments, yielding the situational context used in Figure 5.

A similar approach is possible with compound work problems. Work outputs can also be modeled as *collinear* segments, although their directional orientation is less directly interpretable. In the present analysis, we exclude a sense of direction for work outputs. Working "together" can be modeled as *adjacent* output segments and "competitive" work as *congruent* output segments. For example, the work together (*WT*) problem has *adjacent* output segments that add to yield a single job and *congruent* time segments that, in concert with additive output, allow addition of working rates. This corresponds directly with the situational context of problem *MOD*, without directional orientation of output segments. The competitive work problem (*WC*) can be modeled in a similar fashion. Since Randy and Jo each work on the same set of boxes, we choose *congruent* segments to model the same output. *Adjacent* time segments are associated with the completion of each output, leading to a direct situational correspondence with the round trip problem (*MRT*).

[Insert Figure 6 about here.]

Figure 6 shows a matrix of situational contexts formed by crossing segment relations from output and time dimensions. Compound motion and work problems in each cell have a common situational structure (e.g., problems *MOD* and *WT* in the upper right cell), and off-diagonal cells contain pairs of problems that reverse segment relations for time and output.

For example, reversing *adjacent* distances and *congruent* times in problem *MOD* produces problem *MRT*, provided that *opposite* directions are retained in both problems. Problem structures in diagonal cells of the figure (shaded) are not used in this study but also provide the basis for particular algebra story problems. For example, the lower right cell of Figure 6 contains what Mayer (1981) calls "speed change" problems. This constructive approach to situational contexts can be extended to larger relational vocabularies for output and time (e.g., including *overlap*, *disjoint*, etc.), yielding a sizable space of situational contexts that provide the dimensional basis for algebra "stories" about motion and work.

These examples show that our language of dimensions, places, segments, and inclines can be used to model compound motion and work problems. We have also examined the coverage of this language over different classes of algebra story problems, like those included in Mayer's exhaustive taxonomy (1981). Useful models of situational context can be constructed for most of these classes, including current, mixture, simple interest, cost, and coin problems. Some extensions of the language appear necessary to model relational constraints involving additive and multiplicative comparisons (e.g., "12 more than" or "twice as fast as"). In general, however, models of situational context are possible for any problem in which related linear functions can sensibly be shown within two dimensions. Although arbitrarily complex quantitative relations can be graphed in a Cartesian plane, the provision that their dimensions be "sensible" restricts our modeling language to situations where one-dimensional relations like *adjacent* and two dimensional operators like "joining" or "scaling" have meaning. Thus, dimensional models of situational context may be applicable beyond textbook algebra story problems and include everyday situations involving related linear functions.

Comparison of situational and quantitative structure. Isomorphism within cells and reversed structure across cells of the matrix in Figure 6 partition the space of compound algebra story problems in a way that is complementary to the problem classes described

in the preceding section on quantitative structure. In fact, the problems paired in each cell also have an isomorphic quantitative structure, and problems from off diagonal cells reverse quantitative relations. For example, an additive triad over distance extensives in problem *MOD* contrasts with a shared extensive for distance in problem *MRT*. In our view, this complementarity arises precisely because the quantitative substructures serve as a mathematical abstraction for describing situational contexts. In turn, our relational language of situational contexts provides an abstraction for describing (or modeling) events within particular problems. Thus, choosing segment relations for output and time gives rise to an organized space of situational contexts for compound motion and work problems, each with a corresponding quantitative structure.

While quantitative and situational viewpoints on algebra story problems are complementary, they are not identical. The quantitative network formalism models conceptual entities of time, output, and rate as abstractions that preserve quantitative type (e.g., extensives versus intensives) and value, either as a number or an algebraic expression. In contrast, situational segments and inclines model these same entities as individuals that preserve semantic type (e.g., time versus output), dimensional order (i.e., segments versus inclines), quantitative value, a physical sense of extent (i.e., the length of a segment or the slope of an incline), and local "spatial" relations between individual instances of extent (e.g., the 60 and 100 kilometer segments after the first hour of travel are *adjacent*). Preserving physical extent and relations of locality allows a problem solver to utilize spatial knowledge when identifying or verifving quantitative constraints. For example, when a total distance can be decomposed into component distances which exactly fit within the total, there is a direct physical justification for their addition. "Joining" or "scaling" inclines using a two dimensional model of rate promises a similar physical justification for operations on intensive quantities. Whether students actually use such a vocabulary for justification is an interesting issue, not directly addressed in the present study, that we are exploring

further (Hall, 1987). We suspect that shared situational structure, in addition to quantitative structure, contributes to subjects' judgments of similarity between an arbitrary pair of algebra story problems.

Quantitative and situational structure are not the only materials in the domain of algebra story problems that are important for problem solving, learning, and teaching. Neither can we tacitly assume that these structures, as described above, are actually held by subjects during problem solving. However, these structural abstractions may help to understand what subjects actually do when confronted with a problem to be solved, and to hypothesize what must be learned for competent problem solving to be achieved. Knowledge sources that guide the generation of quantitative representations, and the manner in which they are manifested during problem solving, comprise an important part of competent performance. By grounding quantitative structure in conceptual understanding, these knowledge sources may allow a problem solver to effectively assemble and validate representational structures and operators in the algebraic formalism. Having described some aspects of the underlying situational and quantitative structure of algebra story problems, we now turn to an exploratory study of problem solving.

METHOD

The primary goal of this study *is* to characterize the activities of "competent" problem solvers on representative algebra story problems. When compared with the activities of beginning algebra subjects, the contrast should give a rough image of the terrain over which a learner must travel to become a skilled problem solver. We chose to study subjects who have clearly mastered the algebra curriculum up to existing institutional standards, but who were not recent recipients of algebra based instruction. Thus we are attempting to describe a primary target of traditional instruction in algebra: a problem solver who has mastered the tools of the algebraic formalism, has practiced these skills during instruction, and should be able to apply these skills in novel settings. The study involves minimal

experimental intervention, and our interpretation and analysis of problem solving protocols are primarily descriptive.

Subjects

Subjects in this study were 85 undergraduate computer science majors in their junior and senior years. They were enrolled in an introductory course in artificial intelligence, and participated in the study as part of their classroom activities. These subjects could be viewed as "experts" in algebra story problem solving since they *must have successfully completed* courses in algebra during secondary schooling. In addition, prerequisites to the artificial intelligence course include three university-level courses in calculus and completion or current enrollment in courses covering discrete mathematics. Thus the level of mathematical sophistication in this sample of problem solvers should be high. Alternately, one might argue that these subjects were expert algebra story problem solvers at one time but that their skills have in some sense been "retired" with the passage of time. As will be clear shortly, the solutions offered by many members of this sample do not fit an image of smooth execution of a practiced "skill."

Materials

Subjects were asked to solve the four algebra story problems shown in Table 1. Problems *MOD*, *MRT* and *WT* were taken directly from Mayer's (1981) sample of algebra story problems, with minor alterations in their number set and phrasing. These alterations were intended to free students from unwieldy calculations during problem solving and to make wording between selected pairs of problems more similar. Problem *WC* was constructed to be isomorphic to the *MRT* problem at the level of quantitative structure.

These problems were selected for two reasons. First, with the possible exception of *WC*, they are highly typical of problems found in secondary school texts. Out of an exhaustive set of 1097 algebra story problems drawn from 10 texts, Mayer found that *problems like MOD*,

MRT, and *WT accounted for 7.8% of all observed problems*. Second, different pairings of these problems allow us to present subjects with opportunities for positive or negative transfer across contiguously presented problems.

Specifically, problem pairings *MOD/WT* and *MRT/WC* are isomorphic in their quantitative structure (see Figure 2 for a graphical representation of these pairs) and have similar situational contexts. In the *MOD/WT* pair, output dimensions are *adjacent*, being *collinear* and sharing a starting point, while time dimensions are *congruent*, overlapping completely by sharing both starting and ending times. In the *MRT/WC* pair, outputs are *congruent* while time segments are *adjacent* and of different value (see Figure 3). Should subjects recognize this similarity, they may exhibit some form of positive transfer. Alternately, problem pairings *MOD/MRT* and *WT/WC* are similar at a more superficial level, sharing types of surface materials (e.g., distance traveled or parts of a job completed) while having quite dissimilar quantitative and situational structures. *In fact, relations over output and* time dimensions are exactly reversed, as described in the preceding section on quantitative structure. In the *MOD/MRT* pair for example, outputs in *MOD* are *adjacent* and times are *congruent*, while outputs in *MRT* are *congruent* and times are *adjacent*. When presented contiguously, these problem pairs may induce fairly specific forms of negative transfer (e.g., adding rates in the *MRT* problem after correctly solving the *MOD* problem).

Procedure

Problem materials were distributed so that subjects with adjacent seating during data collection would be in different groups. Group membership was not randomly determined but should reflect no systematic bias. Subjects were allowed eight minutes to solve each problem, and all subjects worked through the problems at the same time. Those finishing early on an individual problem waited until the eight minute time limit expired before proceeding to the next problem. Before solving any problems, subjects were asked to "show all of your work" in a written form, to "work from top to bottom, writing new material

below previous material," and not to erase after making a mistake. Instead, they were asked to mark through any mistake with a single line. Finally, subjects were instructed to "...draw a box around your answer." After solving all four problems, subjects were given 20 minutes to explain their solutions in writing on facing pages of the text booklet without changing their original work.

Problem ordering. The first group of subjects (group M, $n = 46$) saw problems in the following order: *MOD, WT, WC, MRT*. The second group (W, $n = 39$) saw the following order: *WT, MOD, MRT, WC*. Thus, each group solved pairs of problems that were isomorphic at quantitative and situational levels (*MOD/WT* or *WC/MRT*) and also solved pairs of problems that were superficially similar but had reversed relations in quantitative and situational structure (*WT/WC* or *MOD/MRT*).

Data collection. The "behaviors" reported here, and all interpretations of them, are based entirely on subjects' written protocols. Relying solely on written protocols has several obvious disadvantages.

- *There is no timing information.* While students were allowed eight minutes to solve each problem, we can neither determine how long a subject works on any single problem, nor how long any particular written episode lasts — e.g., performing algebraic manipulation.

- *Written material may be a lean or even distorting window on a subject's cognitive processing.* A subject may omit materials that seem unimportant or potentially embarrassing; alternately the subject may give written evidence of processes or strategies that bear little relation to what she actually does.

Since this study is exploratory in nature, we present our results as a heuristic tool for generating hypotheses, and leave more manipulative procedures for confirmatory studies.

Scoring. Written protocols were scored in committee by the authors, using majority rule for categorization of troublesome cases. A scoring system was constructed around the

analysis of problem structure described earlier, using an iterative process with subsets from the total pool of protocols. Scoring categories were added, refined, or dropped from the final system when scorers had persistent difficulty reaching consensus.

THE EPISODIC STRUCTURE OF WRITTEN PROTOCOLS

This section describes a qualitative framework for interpreting written problem solving protocols, showing representative protocols as examples of scored categories within the framework. We point out connections between several of these categories and hypothetical representations and inferences described earlier, although these connections are open to many interpretations. Our framework resembles Schoenfeld's (1985) analysis of mathematical problem solving by concentrating on *coherent episodes of problem solving behavior* (see Ericsson & Simon (1984) for a review of aggregation techniques). We also explicitly score the transition between problem-solving episodes.

A subject's written protocol for a given problem is interpreted in two stages. First the *protocol is divided into a sequence of coherent problem solving episodes*, and then each episode is scored individually with respect to its content, its correctness and its function in the overall sequence. In nearly all cases, the following definition of a problem solving episode allowed scorers to reach consensus:

- *Strategic coherence.* The subject is pursuing the same overall goal.

- *Tactical coherence.* The subject is using the same method for attaining this goal.

- *Conceptual coherence.* The subject is exhibiting the same conceptualization of the problem.

Although episodes divide problem solving into coherent chunks, the context created by earlier episodes is assumed to be inherited by later ones, unless there is evidence that a reconceptualization has occurred. Our definition of an episode will be sharpened in the

following paragraphs as we specify in detail the scoring categories used to describe episodic content.

After dividing the written protocol into coherent problem-solving episodes, each episode is examined to determine its general content. Content categories include: strategic purpose, tactical content, conceptual content, the presence of conceptual or manipulative errors, and finally the status of the episode in the overall solution attempt. The latter covers relative correctness and the reason for transition to a new episode. With the exception of conceptual content, each of these categories is further differentiated into alternative subcategories, as shown in Table 2. In some cases only one subcategory is selected as best describing the more general category (e.g., simulation as a type of model-based reasoning under tactical content); in other cases, each subcategory can occur within a single episode (e.g., various kinds of conceptual and manipulative errors).

[Insert Table 2 about here.]

The remainder of this section takes up each of these interpretive categories in detail, showing representative written protocols as examples of their use in scoring the episodic structure of subjects' solution attempts. For example, subject m20 in Figure 7 goes through three error-free episodes, each with a specific purpose, tactic, content, and transition. In the protocols shown in figures as illustrations of various categories, episodes are separated by dashed lines, and their sequence is shown with circled numbers. Several protocol excerpts are presented directly in the text without accompanying figures.

[Insert Figure 7 about here.]

Strategic purpose

The strategic purpose of an episode is its relation to the ultimate goal of finding a solution. Judgments of a problem solver's "purpose" are clearly a matter of our own interpretation, although we present scoring criteria that make these judgments operational across individual ratings. In this regard, our scoring distinguishes between three abstract problem solving

modes.

Comprehension. The subject is not directly seeking a final solution, but is constructing a representation of the problem by incorporating various constraints. In episode 1 of Figure 7, the subject finds a way to express working rates by considering their outputs after one hour.

Solution attempt. The subject is attempting a series of operations that work directly toward a solution *(Figure 7, episode 2)*.

Verification. The subject has already produced a solution to the problem and is now seeking confirmatory evidence for it, for instance by rederiving the solution with another method or by inserting the answer in some intermediate equations (Figure 7, episode 3).

Tactical content

The tactical content of an episode is the method used by a subject to achieve some strategic purpose. Our operational criteria refer primarily to the protocol material for the current episode, but in a few cases information contained directly in the protocol was insufficient to make an operational category judgment. In these cases, surrounding episodes and *post hoc* written explanations supplied by the subject were used to assist scoring.

[Insert Figure 8 about here.]

Annotation. These episodes usually occur early in the protocol when subjects are collecting information about the problem. Three cases are covered.

- *Problem elements.* The subject is recording elements of the problem text (e.g., $V_A = 60km/hr$, Figure 8, episode 2).
- *Retrieval of formulas.* The subject is remembering and writing down memorized formulas which seem relevant, (e.g., $v = \frac{d}{t}$, Figure 8, episode 4).
- *Diagram.* The subject draws a pictorial representation of the problem situation (e.g., Figure 8, episode 1).

Algebra. An episode is algebraic if it makes use of one or more equations placing constraints on the value of one or more variables. However, simple assignments are not treated as equations. Thus neither $100 + 60 = 160$ nor $d = 880$ are considered equations, while $d = 100 \times t$ is considered an equation. As shown unusually clearly in the protocol of Figure 9, the tactical approach of the typical algebraist is to express constraints as a system of one or more equations (or proportions) and to solve for the appropriate unknown. We have also found cases of subjects trying equations in a generate-and-test fashion until, as one subject explained, an equation "looks good."

[Insert Figure 9 about here.]

Model–based reasoning. This category is scored when a subject "executes" a model of the problem situation along the dimension defined by an unknown quantity such as time, distance or work. Subcategories of model-based reasoning relate to constructive problem solving inferences described in the preceding section on situational structure.

- *Simulation*[7]. The subject selects a base unit for the chosen dimension and "runs" the model for each successive unit increment as illustrated in episode 3 of Figure 8. Consistent with our earlier development of situational structure, a simulation episode could be interpreted as an iterative "joining" of concrete individual inclines. Simulation can also be partial (just one or two increments) in that it is not used to reach a solution, but to examine relations between quantities and to enable some other solution method. In both episode 1 of Figure 7 and episode 5 of Figure 13, a simulation for one hour establishes the quantitative combination of entities from distinct events.

- *Heuristic*. The base quantity "jumps" by variable increments whose magnitude is determined at each point by estimations of closeness to the solution. A heuristic model based reasoning episode could be interpreted as "scaling" inclines that represent invariant relations, as described earlier. The progression of this generate and test approach can be monotonic, as in episode 2 of Figure 10, or follow some form of interpolation search. After each generation of a value, the state of the problem situation being modeled is reconstructed and evaluated.

[Insert Figure 10 about here.]

Ratio. This subcategory covers a number of tactics by which relations of proportionality between quantities are used, sometimes providing clever "shortcuts" to a solution. These tactics clearly utilize a representation of quantity (e.g., intensive quantities, as described earlier), but the manner in which related quantities are integrated may depend upon constructive inferences within the situational context (e.g., composing segments or inclines).

[Insert Figure 11 about here.]

- *Whole/part*. The subject views a part as fitting some number of times into a whole quantity, as in episode 6 of Figure 13.

- *Part/whole and part/part*. These two types of ratios compare portions of entities. Use of the part/whole ratio is illustrated in episodes 2–4 of Figure 11, where the subject considers parts of the total job[8]. A version of the part/part ratio appears in episode 2 of Figure 12, involving the respective rates of bus and foot travel.

- *Proportion*. Non-algebraic proportions cover reasoning of the type exhibited by subject m05 on the work together (WT) problem: "... they've done $\frac{9}{10}$ [of a job] in 2 hrs, so $\frac{2}{9}$ hr more would do for [the job] left to be done ..."

[7] Our use of "simulation" is somewhat different from its use in computational studies of common sense reasoning. For example, de Kleer's (1977, 1979) "envisionment" uses quantitative calculation to resolve qualitative ambiguity, while our sense of simulation uses physical construction to help disambiguate quantitative constraints.

[8] Although this protocol illustrates the category clearly, it is probable that successful use of this ratio was somewhat fortuitous on the part of this student, since a general justification for its validity is rather complex.

- *Scaling.* The subject solves a related version of the problem or reaches an unexpected answer, and simply scales the answer to fit the quantities given in the problem. This may relate to our earlier description of "scaling" rates as invariant two dimensional inclines. In episodes 3 4 of Figure 12, for example, the subject solves an easier problem by *heuristic model-based reasoning* and then scales her answer to "fit" the *MRT* problem.

[Insert Figure 12 about here.]

Unit. In a few cases, a subject reasons purely in terms of *units of measurement given in* the problem. For instance, on the work competitive problem (WC), subject m44 examines alternative rate forms with the following manipulations:

$$\frac{box}{min(utes)} \cdot min = box \qquad \frac{min}{box} \cdot box = min$$

Procedure. This subcategory is scored when there is evidence that a subject is executing some stored sequence of actions or operations other than routine algebraic or arithmetic manipulation. For example, on the work together problem (WT) subj 21 appears to use a simple averaging tactic for combining quantities, writing "$total = \frac{1}{2}(5+4) = \frac{9}{2} = 4\frac{1}{2}hrs.$"

Conceptual content

The conceptual content of an episode reflects the subject's conceptualization of the problem situation and the resulting set of constraints between problem entities. There is a subtle *but crucial distinction between situational understanding* and the quantitative constraints that are implied by it, as suggested in previous sections. Without further subcategorization, our scoring of conceptual content simply contains the constraints that the subject clearly recognizes and uses in the episode. For instance, subject m39 in Figure 9 manifests an understanding of all necessary constraints: equal distances, additive composition of times, and the distance rate time relation.

[Insert Figure 13 about here.]

Errors

Within each problem solving episode, we consider two broad classes of errors.

Conceptual errors. These are scored when a subject either includes a constraint that is inappropriate for the problem or excludes a constraint that is a critical requirement for the current episode.

- *Errors of commission.* These errors are incorrect constraints that the subject introduces during an episode, either by incorrectly representing the situational context of the problem or by making erroneous quantitative inferences. For example, in episodes 4 6 of Figure 13 the subject subtracts distances because she thinks that the trains are going in the same direction.

- *Errors of omission.* These errors are overlooked constraints. To be scored as an error of omission, an overlooked constraint has to be critical to the solution method applied by the subject. This usually means that two entities are explicitly used while the relation between them is ignored. In Figure 14, episode 3, the subject has overlooked that working times represented as x and y are equal.

[Insert Figure 14 about here.]

Manipulation errors. Since written protocols usually display algebraic or arithmetic manipulations clearly, our scoring identifies manipulative errors of three types.

- *Algebraic errors.* For example, on the MOD problem, subject w39 writes "$880 = \frac{160}{T}$" followed by "$T = \frac{880}{160}$."

- *Variable errors.* We observed two types of errors related to the concept of variable. In "switch errors," the meaning of a variable changes in the course of problem solving. In "label errors," subjects are using variables as labels for quantities. For instance, in

the round trip problem (*MRT*), subject m10 writes the equation "$1B + 8W = 6hrs$" and explains that "for every 1 hour on the bus, it takes 8 hours to get back."

- *Arithmetic errors.* For example, on the opposite direction motion problem (MOD) subject m20 writes "[illegible] = $\frac{11}{4}$." After detecting this arithmetic error in a verification episode, the subject recovers by using the ratio scaling tactic mentioned earlier.

Status of episode within solution attempt

Categories listed so far deal with internal characteristics of an episode. The two aspects of the scoring scheme described here, consistency and transition, concentrate on the relation *of an individual episode to the overall problem-solving effort.*

Consistency. This category assesses the correctness of an episode in the context of the problem-solving sequence and is scored correct or incorrect for three facets.

- *Before.* This subcategory reflects the correctness of the context inherited by the episode. For example, errors may be generated in *former episodes* and passed into the current episode, as with the conceptual error of commission (same direction) passed between episodes 4 and 5 of Figure 13.

- *During.* This scores the correctness of the current episode with respect to the inherited context. An episode producing an incorrect result can be internally correct if it is consistent with an incorrect context. For example, episodes 5 and 6 of Figure 13 are internally consistent with the conceptual error of commission introduced in episode 4.

- *After.* This subcategory assesses the absolute correctness of the outcome of the current episode. If a solution is presented, the scoring reflects its correctness, otherwise scoring assesses whether or not the subject is on a possible right track.

Transition. The intent here is to determine the reason why a subject passes from one episode to the next. Unlike consistency, which reflects the scorers' judgment of correctness, this aspect attempts to capture the subject's point of view.

- *Subgoal.* The subject accomplishes an intermediate goal, bringing the episode to an end (Figure 7, episodes 1 and 3). Information identified when achieving a subgoal (e.g., changing the form of a working rate) is generally carried into subsequent episodes.

- *Wrong.* The subject decides that she is on the wrong track and abandons the current approach, usually by marking through the work (Figure 13, episode 3). This transition is often the result of an explicit verification episode.

- *Impasse.* The subject reaches a point where she cannot continue with the current method. A good example of impasse is shown in episode 3 of Figure 8, where the subject correctly applies simulation by hourly increments, overshoots the non integer solution, and then switches to an algebraic tactic after adding rates.

- *Lost.* The subject reaches a point where she cannot determine how to proceed, as in episode 2 of Figure 14.

- *Final solution.* The subject reaches a result and presents it as a solution to the problem.

- *Found solution wrong.* The subject realizes or believes that the solution presented is incorrect.

This presentation of our framework for interpreting written protocols gives an overly linear picture of its use in scoring subjects' solution attempts. In fact, categorizing the episodic structure of a written protocol within this framework was usually done quickly (from 5 to 20 minutes per protocol) and with little subsequent disagreement among the

scorers. By design, each category was rated with at least 75% agreement over four scorers; most categories approached unanimous agreement.

In addition to determining whether or not a subject has managed to find the "correct" solution to an algebra story problem, this framework for interpreting problem solving episodes allows us to describe the internal structure of the subject's solution attempt. Our interpretation of episodic structure supports more fine-grained *explorations* of the strategic and tactical course of problem solving. In the quantitative results section that follows, we form composite analytic categories by identifying episodic patterns among the atomic category judgments described above. Thus we will be able to speak of subjects reaching a "final episode" with some particular tactic and content or to examine a series of contiguous episodes during which model-based reasoning is used. Beyond the results presented here, we expect the set of scored protocols to provide a rich dataset for continuing analysis.

QUANTITATIVE ANALYSIS OF PROBLEM SOLVING EPISODES

In the section on problem structure, we argued that competent problem solving proceeds as an elaborative, interdependent exploration of two distinct problem spaces: the situational context of a story problem and the quantitative constraints given explicitly or implied in the problem statement. Results presented *in this section* provide evidence for this interdependency at a global level of problem-solving activity and at a more detailed level of episodic content. Our analysis distinguishes between *subjects' problem-solving* attempts and the episodic structure of those attempts. By *problem-solving attempt*, we mean all of the activities evident in the written protocol, which may include several distinct episodes. By *episodic structure* we mean the alternation of problem-solving episodes of various types, and the constraints or errors that are contained within and across those episodes.

First we examine the tactical content, strategic purpose, *transitional status, and errors* present in *subjects' solution attempts*. These analyses pool episodes within solution attempts to show the prevalence of different interpretive categories, and so they provide

only a coarse image of competent problem solving. Second, we look within individual solution attempts and examine two episodic patterns in detail. *An analysis of the episode* during which a *final solution is offered* provides a finer image of problem solving outcome, describing relations between solution outcomes and other interpretive categories within the episode. We also identify individual episodes of *model based reasoning* to permit a closer examination of problem-solving activity outside of the traditional algebraic formalism. By considering the content of surrounding problem solving episodes, we can begin to examine *subjects' reasons for* using model based reasoning and to assess its effectiveness for making correct problem-solving inferences or recovering from existing errors. The section ends with a summary of major quantitative findings.

Problem-solving attempts

Since many of our rated categories represent hypotheses about problem solving processes, we present their frequency of occurrence within subjects' problem solving attempts. Table 3 shows the percentage of subjects having one or more episodes in which various rated categories were observed. Percentages are shown separately for each problem (*MOD, MRT, WT, WC*) but are collapsed over groups (*M, W*) since none of these contrasts were statistically reliable. Most findings are as expected, while several are surprising.

[Insert Table 3 about here.]

Tactical content of scored episodes. While most subjects use algebra in their solution attempts (63.5 to 85.9% across problems), reasoning within the situational context presented by the problem is surprisingly common.

- Looking within individual problems, at least one *model based episode* is used by 22.4% to 47.1% of subjects, depending on the problem. A separate analysis pooling across problems shows that 72.9% of subjects have one or more episodes of model based reasoning in their written protocols. These episodes are explored more fully later.

- Use of ratios is the next most prevalent non-algebraic tactic (14.1% to 42.4% across problems) and may depend upon a variety of factors: the complexity of the constraints presented by a problem's quantitative structure, the accessibility of situational justifications for those constraints, and the manner in which the constraints are presented in the problem text.

- Few solution attempts contain episodes using a "procedure" or reasoning with "units." Most subjects using a procedure on the *WT* problem chose to take an average over working rates, a strategy that violated the situational meaning of "working together" in that problem and generally led to an incorrect solution.

- Annotations, in the form of diagrams or notations about problem elements, were either scarce or common, depending upon the situational and surface content of the story problem. Motion problems (*MOD, MRT*) showed few notations (7.1%, 15.3%) but more frequent diagrams (69.4%, 36.5%), while work problems showed frequent notations (21.2%, 29.4%) but fewer diagrams (8.2%, 9.4%). Although it is likely that the spatial content of motion problems makes them more accessible to diagrammatic representation, some subjects are able to construct effective diagrams for work problems (e.g., see Figure 11, episode 3).

Strategic purpose of scored episodes.

- Most subjects show explicit attempts at comprehension in their written protocols (57.6% to 84.7% across problems), typically through diagrams, notations or model based reasoning.

- While all subjects make some attempt to solve the problem, only a minority give evidence of attempting to verify the results of their work (7.1% to 28.2% across problems).

Transitions out of scored episodes.

- Most subjects find and explicitly present a solution (either correct or incorrect) as part of their problem solving attempt, although problems *MRT* and *WT* appear more difficult than their quantitative isomorphs in this regard (*WC* and *MOD*). A more detailed analysis of solution outcomes follows shortly.

- Likewise, the three transitions without solution (i.e., impasse, lost, or wrong) are most common in the more difficult problems (*MRT* and *WT*).

Errors in scored episodes.

- Conceptual errors of omission and commission increase for the more difficult problems (*MRT* and *WT*), and appear much more frequently than manipulative errors (arithmetic, algebraic, or variable errors) on all problems.

Several interesting patterns emerge in these findings. First, subjects' written protocols are not composed solely of material generated while performing algebraic transformations. Instead, many subjects appear to use various forms of direct situational reasoning, which we have termed *model based reasoning*, conducted within their understanding of the context posed by a story problem text. Second, although most subjects do present a solution in some form, their efforts do not appear as a smooth progression toward a quantitative solution. Rather, their problem-solving efforts are often interrupted by varied conceptual difficulties that must be repaired before a solution is found. Third, manipulation errors within algebraic and arithmetic formalisms do occur, but these are overshadowed by conceptual errors of omission or commission as a primary source of problem solving difficulty. Consistent with our earlier treatment of problem structure, we interpret these findings to mean that students form an understanding of the problem at the level of its situational context and then use this understanding to introduce quantitative constraints. As a result, many of the activities

present in an episodic analysis of algebra story problem solving fall outside the traditional algebraic formalism.

Final episodes: outcome, tactical content, and errors

Examination of the written protocols clearly shows that subjects undertake a variety of problem-solving activities when attempting to solve these problems, particularly when they encounter difficulties in reaching a solution. However, the previous findings speak only to the *presence* of various conditions in subject's problem-solving efforts. By our scoring, subjects averaged approximately 2.5 scored episodes per problem-solving effort, with some protocols presenting evidence for as many as 10 distinct episodes. In the following analyses, we look within individual protocols for more finely-detailed episodic structure.

Within an individual's efforts on any given problem, we extract a "final episode" for a first level of detailed analysis. This episode need not be the subject's last effort in a solution attempt, but it is final in one of three senses: it is the last episode during which a subject presents a solution that is correct, the last episode during which they present a solution that is incorrect, or the last episode of a problem-solving effort in which no solution is presented. "Incorrect" means the subject presents an incorrect final solution without detecting any errors. The "no solution" category includes subjects who present an incorrect solution but realize they have done so during a subsequent attempt at verification, without being able to recover. Thus, the final episode may be either correct, incorrect, or present no solution.

[Insert Table 4 about here.]

Performance outcomes across groups. Table 4 shows the final outcomes for each problem, broken out to show anticipated effects of problem ordering. For example, on problem MOD group W should perform better than group M (shown as M < W in the table), since subjects in group W are exposed to an isomorphic problem (WT) just before seeing problem MOD. If positive transfer occurs, subjects in group M should be at a relative disadvantage, having seen no prior problem. None of the group contrasts were statistically

significant, even taking into account whether subjects were correct or incorrect on preceding problems. Thus, the problem ordering manipulation introduced to provide opportunities for positive and negative transfer appears to have had little effect on subjects' performance at the level of solution correctness. We consider this finding at a more detailed level in the discussion section. Clearly, problems *MRT* and *WT* were most difficult, with percentages of subjects reaching a correct solution on these problems (51.8% and 61.2%) falling well below those reaching correct solutions on problems *MOD* and *WC* (90.6% and 91.8%).

[Insert Table 5 about here.]

Relations between solution outcome and tactical content. Table 5 shows tactical content and error categories for final problem solving episodes. For tactical content, a subject receives a single category score, so cell frequencies sum to give appropriate column totals. A few protocols contain insufficient information to score tactical content in the final episode. For errors, a subject may achieve a correct solution in the final episode but still demonstrate an error, or they may have several types of errors. As a result, cell entries for errors do not always add up to coincide with column totals.

The prevalence of tactical content and error categories in the final episode is generally consistent with findings for overall solution attempts. However, by looking within these attempts we can focus more closely on relations between tactic and outcome.

- Even within the final episode, not all solutions (correct or incorrect) are found using algebra. Excluding those with no solution or with contents that were not scorable, between 22.0% and 44.0% of subjects (across problems) used other tactics to find their final solution.

- Use of ratios is the most prevalent form of non algebraic reasoning in final episodes, with the exception of an incorrect averaging procedure on problem *WT*. Model based reasoning is the next most prevalent tactic.

- Algebra, model based reasoning, and ratio tactics are about equally effective in the

final episode. Pooling across problems, algebra is slightly more successful (number correct/total observed) and slightly less error prone (number incorrect/total observed) than either of the non algebraic tactics.

Thus, even within the final episode where a solution might be found, a normative account of problem solving consisting of successive algebraic transformations would be disconfirmed by these data. Instead, subjects find solutions through a *variety of reasoning strategies* that, in some cases, involve *relatively little formal algebra*. In a moment, we examine the *episodic structure of model-based reasoning tactics* more closely.

Relations between solution outcome and errors. Errors observed during final episodes are also interesting although more difficult to interpret since individual subjects can have multiple errors. We distinguish between "*conceptual errors*," which arise through omission or commission of specific quantitative constraints, and "manipulative errors," which arise through improper use of arithmetic, algebraic operations, or variables. These error categories are shown in the lower panel of Table 5.

- With the exception of problem *MOD*, conceptual errors are more prevalent than manipulation errors. This is particularly true of the more difficult problems (*MRT* and *WT*).

- Subjects who achieve a correct solution have fewer conceptual errors than those with an *incorrect solution or no solution* (1:6, 0:30, 1:37 and 1:4 across problems). In the few cases where a solution is found despite conceptual errors, offsetting manipulative errors fortuitously "correct" these conceptual errors.

- Although *manipulative errors* are found among subjects who do not reach a correct solution, they are also observed among subjects giving a correct solution. These errors are repaired within the final episode to allow for a correct solution.

- Among subjects who reach an incorrect solution, the number with *manipulative errors* could not account *for more than a third of these* failures (2/6, 5/15, 7/21, and 1/5 across problems). Alternately, at least two thirds of the incorrect solutions must be based on conceptual errors.

One interpretation of these results is that manipulative *errors are less frequent and more recoverable* than *conceptual errors*. That is, subjects who make an error during a problem solving episode are more likely to recover from that error if it stems from arithmetic or algebraic manipulation than if it is a result of misunderstanding or misencoding the structure of the problem. Since errors may *persist across episodes*, this conclusion cannot be unambiguously supported. Nonetheless, the most serious errors among this group of relatively competent problem solvers are conceptual rather than manipulative.

Episodic structure of model-based reasoning

One of the most intriguing findings in these data are subjects' use of what we call "model based reasoning." In these episodes, subjects depart from the algebraic formalism and reason directly within the situational context presented by the story problem. In this section, we examine the functional *role* that model based reasoning plays within the overall solution effort. We are interested in determining under what circumstances this form of reasoning occurs, what purpose it serves within a particular solution attempt, and what outcomes are likely when subjects reason in this fashion.

As with the analysis of final episodes, we identify specific episodes within subjects' solution attempts where model based reasoning occurs. We also extract the preceding problem solving episode in the hopes of identifying enabling conditions for model based reasoning. Since some subjects' only use of model based reasoning occurs *during their first* scored episode, they will have no preceding episode.

[Insert *Table 6 about here*.]

Precursors to model–based reasoning. *A first task for describing the role of model based reasoning in subjects' solution attempts is to determine their reasons for using this method. We will contrast the correctness and transition out of an immediately preceding episode with the purpose (as we have rated it) for using model based reasoning.*

Table 6 shows the number of subjects who use model- based reasoning for some purpose (scored as comprehension, solution attempt, or verification) subsequent to various conditions in the preceding episode. A subject may either have no preceding episode, have a preceding episode without errors, or have a preceding episode with one or more scored errors (i.e., an error of commission, omission, or manipulation from which the subject does not recover in that episode).

- From 26.3% (5 of 19 on *MRT*) to 70.0% (21 of 30 on *WT*) of model- based reasoning episodes occur as the first episode in a solution attempt.
- Of these initial model-based episodes, the majority (except for problem *MRT*) are undertaken for the apparent purpose of comprehending some aspect of the presented problem. The remaining initial episodes are scored as solution attempts.

For subjects having a preceding episode, their transition out of this episode is scored as achieving a subgoal, finding a solution, reaching an impasse, or deciding they are wrong. Of the model-based reasoning episodes following an error-free episode, there are two essentially different conditions. In the first, a subject's preceding episode ends with achieving a subgoal or finding a solution. This subject could be considered "on track" in her solution attempt. In the second condition, subjects "abandon" the preceding episode after reaching an impasse (also after getting lost, as described earlier) or deciding that their efforts are wrong. These subjects are technically on track since their preceding episodes are free of errors, but they encounter sufficient difficulty that they abandon a previous line of *reasoning in favor of* model based reasoning.

- Amost all subjects who are "on track" in a preceding episode either attempt a solution or continue attempts at comprehension during the model based reasoning episode.
- Only a few subjects are "on track" and undertake model based reasoning for the purpose of verification. On problem *WC* these verification episodes follow finding a solution; the single verification attempt *on problem WT comes from a subject who* verifies a recalled formula using a simplification of the original problem.
- Subjects "abandon" (i.e., lost, impasse or wrong) a prior, error free episode infrequently and subsequently use model based reasoning for comprehension or to attempt a solution.

Model based reasoning episodes following an episode with errors are less frequent than those discussed above, but fall into similar categories. Relatively few subjects have preceding errors, are unaware of those errors, and proceed as if "on track" (achieve a subgoal or find a solution). Subjects who are aware of their preceding error nearly always decide that they are wrong and "abandon" the preceding episode.

- Among those who "abandon" a preceding episode with errors, subsequent model based reasoning is used either for comprehension or as an *attempt to find a solution.*

Although based on a subset of all subjects studied, these findings support an interpretation in which model based reasoning plays four basic roles in problem solving: as a *preparatory comprehension* strategy when the model based episode is either the first problem-solving activity attempted or follows other comprehension episodes, as a *solution* strategy when subjects feel they are on track, as an *evidence gathering* strategy when a solution has been found previously (this is infrequent), or as a *recovery* strategy when subjects suspect that their comprehension or solution efforts may be "off track." These interpretations are [illegible]ent with our earlier argument that reasoning within the situational context

of a problem supports the generation of quantitative constraints, can be used directly as a solution method, or can be used to verify that these constraints are appropriate.

[Insert Table 7 about here.]

Effectiveness of model–based reasoning. As well as inferring subjects' reasons for undertaking model-based reasoning, we would like to characterize the effectiveness of this reasoning strategy. To assess efficacy, we examine the occurrence of any errors within successive episodes. Table 7 shows the relationship between errors during a preceding episode (when there is one) and errors within the model-based reasoning episode.

- When model–based reasoning is the subject's first evident activity, as indicated by "No episode" in Table 7, errors are not often encountered within that episode. The two errors shown for problem *MRT* are mis-conceptualizations in which subjects assume that round trip times are equal. The error in problem *WT* comes from a subject who assumes that Mary and Jane do equal amounts of work.

- When a previous episode contains errors, the subsequent model–based episode is usually error–free. Thus, existing errors may be "repaired" during model-based reasoning.

- Following an error-free episode, only one subject introduces a new error with model-based reasoning by omitting the constraint that distances are equal on problem MRT.

While these findings are not conclusive, they are again consistent with the four hypothetical roles for model-based reasoning described in the analysis of final episodes. Preparatory comprehension promotes an error-free conceptualization of the problem situation, enabling subjects to correctly assemble the quantitative structure of the problem during later reasoning episodes. Subjects also attempt to find solutions directly through model based reasoning, generally without introducing errors. Alternately, after encountering an error during previous problem-solving activities, subjects may be able to recover through the

use of model based reasoning. Finally, model based reasoning can play a confirmatory role when subjects have identified important problem constraints or a possible solution.

Summary of quantitative findings

As part of our effort to explore the episodic structure of algebra story problem solving, this section presents three levels of quantitative analysis: the prevalence of different interpretive categories in subjects' overall *solution attempts*, relations between outcomes, tactical content, and errors in subjects' *final episodes* of problem solving, and the role and effectiveness of *model-based reasoning* episodes within the wider problem solving context. Each successive level of analysis tightens the focus on findings at coarser levels.

A global view of solution attempts reveals significant non algebraic reasoning as a prevalent and somewhat unexpected constituent of competent problem solving. Most prevalent among these tactics is model based reasoning. Among observed errors, conceptual omissions or commissions are more frequent than manipulative errors within arithmetic or algebraic formalisms. An examination of final episodes, the "bottom line" in a very lean view of these problems, corroborates this global image of significant non algebraic reasoning on non-routine problems. Looking more closely at errors, we find that manipulative errors are both less frequent and less damaging than conceptual errors, since subjects are more likely to recover from errors of manipulation within the final episode. Finally, we examine the episodic structure of model based reasoning and propose four roles for this tactic: as preparatory comprehension, as a solution method, as evidence gathering for a candidate solution, or as a recovery method for errors generated earlier in the solution attempt. These quantitative analyses of problem solving agree with out earlier description of the interplay between the quantitative and situational structure of algebra story problems.

DISCUSSION

Interpreted as a series of problem solving episodes, the written protocols described

above provide an opportunity to look within individual solution attempts for evidence of strategic and tactical approach. We have also been able to look across a relatively large sample of mathematically sophisticated subjects in an effort to describe "typical" problem solving behaviors. In this section, we compare the results of our study with other research on mathematical problem solving and discuss the implications of these findings for conceptions of mathematical "knowledge" and instruction.

Competent problem solving

Our findings are offered as a preliminary exploration of "competent" algebra story problem solving. By choosing the term competent, we hope to contrast the problem-solving behaviors we have observed against images of "expertise" in problem solving as they are often portrayed in the literature. For example, Hinsley *et al.* (1977) and Mayer *et al.* (1984) report that experienced problem solvers use problem-solving schemata to categorize problems by type and then represent these problems using familiar quantitative constraints. While this account corresponds with some of our protocols, many subjects in our sample appear to *construct* solutions to algebra story problems. Rather than a smooth execution of a highly practiced skill, these constructions often proceed with some difficulty and include reasoning activities only partly connected to algebraic or arithmetic formalisms.

As noted earlier, subjects in this study should be considered mathematically sophisticated. Nonetheless, judging from the varied behaviors we have observed, the algebra story problems we presented to subjects are not routine problems. On problems *MRT* and *WT*, for example, many subjects fail to reach a correct solution, and those who do succeed often experience considerable difficulty. Analyses of errors encountered by subjects when attempting solutions suggest that conceptual errors of omission and commission are both more prevalent and more damaging than manipulative errors in algebra or arithmetic. These results support a model of algebra story problem solving in which problem comprehension and solution are complimentary processes. Integrating dual representations of a

problem at situational and quantitative levels is a central aspect of competence. These intermediary structures provide a representational bridge between the text of a problem and a quantitative solution. Reasoning about the situational context of a problem can serve as a justification for assembling quantitative constraints that may eventually lead to a correct solution. Thus, a substantial portion of a subject's activity is devoted to *reaching* an understanding of the problem that is sufficient for applying the routine of formal manipulation.

Despite their mathematical backgrounds, perhaps our subjects have yet to achieve competent algebra story problem solving, well beyond the curricular setting designed to teach it. Alternately, they may have been "experts" during and shortly after algebra instruction, but with the passage of time have lost the facile performance demonstrated by Hinsley *et al.* (1977). Whichever explanation is chosen, the issue remains how to characterize ostensibly competent problem solving in a population for whom the algebra curriculum is designed. Recent studies of mathematical problem solving in "practice" (Carraher, Carraher, & Schliemann, 1987; Carraher & Schliemann, 1987; and de la Rocha, 1986) present similar images of competent quantitative reasoning: problem solvers organize their quantitative knowledge around the demands of the situational context presented by the task, often using the problem situation (or knowledge of it) to assemble or verify quantitative constraints. If an image of competent problem solving in this domain is to inform teaching efforts — i.e., it is to have some predictive capacity as described in the introduction of this paper — then activities like these are a legitimate topic of study. We return to issues of competence and acceptable transitional performance in a moment.

Transfer effects

Aside from their use as representative problem solving tasks, *algebra story problems* often serve as materials for studies of analogical transfer. Given a target problem to solve, subjects exhibit positive transfer when they can use the solution method from a previ

ously encountered source problem to help solve the target problem. Alternately, subjects exhibit negative transfer when they access and use the solution from an inappropriately related source problem. Studies of analogical transfer with algebra story problems have produced mixed results, but show that both positive and negative transfer sometimes occur. Positive transfer has been more likely when subjects are alerted to the experimental manipulation (Reed, 1987; Reed, Dempster, & Ettinger, 1985) or are high in mathematical achievement (Novick, 1987). Transfer effects related to higher achievement have been attributed to subjects' improved attention to aspects of quantitative structure (Novick, 1987; Silver, 1979) and better memory for previous solution methods (Silver, 1981). Negative transfer in subjects with lower achievement (Novick, 1987) has been attributed to a reliance on inappropriate problem features and an inability to reject misleading analogical sources. Finally, Dellarosa (1985) has experimentally manipulated subjects' use of analogical and schematic problem comparisons to produce improvements in their categorization and solution of related problems.

In the present study, we did not alert subjects to the comparability of problems, nor did we encourage them to look back over their prior solutions as they worked through the problems. Their backgrounds insure high mathematical achievement, and entrance requirements for academic majors in computer science and engineering preselect for high quantitative abilities. There is no performance-level evidence of positive or negative transfer within the problem-solving session, despite our manipulation of problem structure and presentation order to elicit these effects. At the aggregate level, our subjects appear to take the "school math" task we present them at face value: each problem, presented individually on a blank sheet of paper, is treated as a self contained exercise, rather like what a student might face during examinations in a course on algebra. However, on closer inspection of individual protocols and explanatory remarks we find that several subjects give evidence for some form of negative transfer.

In some cases, transferred material directly violates the quantitative and situational structure of the target problem. For example, subject w08 incorrectly attempts to add working rates on problem *WC*, first writing $1/5 * boxes + 1/2 * boxes = 56$, followed by $7/10 * boxes = 56$. In explanatory remarks, w08 states that "The mistake I made was that I assumed it was like problem 1 where they work together." In the preceding solution to *WT*, this subject had written "Together $= 1/5 + 1/4$ in one hour $= 9/20$" and then correctly divided one job by the combined rate. Adding working rates in problem *WT* is justified since Mary and Jane work together at the same time. However, situational and quantitative relations are exactly reversed in problem *WC* (see Figures 6 and 2(b)). Since times are added together (*adjacent*) and work is performed on the same boxes (*congruent*), the addition of working rates (i.e., output over time) cannot be similarly justified.

In other cases, subjects recognize an appropriate source problem, but then fail to transfer information at the correct level of abstraction. For example, on problem *MOD* subject w01 correctly attempts to add motion rates, but uses an algebraic expression of the form: $1/60 + 1/100 = x/880$. On the previous (*WT*) problem, the subject manages a correct solution using an expression of the form, $1/5 + 1/4 = 1/x$, and remarks that this "... is a formula used to find a total of time they work together." Although the addition of rates can be justified in both problems, it appears that the rate form in the retrieved formula is reversed (i.e., time over output) when used in a solution attempt on the *MOD* problem. Thus in a situation where we anticipate that the subject will benefit by transfer of a solution approach, their failure to justify transferred material actually produces a negative effect.

It may be that the problem solving context, completing a test booklet in a proctored examination setting, as well as our decision not to alert subjects to the comparability of problems, prevented them from recognizing and elaborating effective analogical comparisons between problems. In more detailed verbal protocol studies where subjects are encouraged to make problem comparisons (Hall, 1987, 1988), attempts at analogical inferences between

algebra story problems are quite common. These comparisons are usually lengthy and can introduce misconceptions, but also frequently lead to fruitful explorations of problem structure, both quantitative and situational. In addition, comparisons need not encompass the entire problem structure, but often instead make effective use of relevant substructural similarities. These alternative findings are largely consistent with other verbal protocol studies of learning from worked examples (Pirolli & Anderson, 1985; Singley, 1986; Chi, Bassok, Lewis, Reiman, & Glaser, 1987), and suggest that analogical comparison may be a common problem-solving and learning strategy in settings where subjects have some control over their work.

Model–based reasoning

We are not the only researchers to note the prevalence of model-based reasoning during mathematical problem solving. A number of psychological studies have found similar evidence, although interpretations of this activity vary. Paige & Simon (1966), comparing human protocols with Bobrow's (1964) computational model of translating algebra story problems into equations, found that subjects with varied mathematical backgrounds used "auxiliary representations" of the physical setting of a problem. These representations allowed some subjects to detect impossible problems or to assemble relevant quantitative constraints (e.g., additivity in part-whole relations). Using verbal protocols to study the prevalence of Polya's (1945) heuristics for mathematical problem solving, Kilpatrick (1967) reported that 60% of an above-average group of eighth graders used "successive approximation" while attempting to solve word problems. These trial-and-error approaches were often successful and were sometimes combined effectively with more deductive solution strategies. Silver (1979) found similar successful approximation strategies in students who had yet to receive formal algebraic training. Studying geometry problems, Schoenfeld (1985) found that students used a trial-and-error approach to generate hypotheses about geometric relations and then evaluated these hypotheses by physical construction. He argued that

these exploratory episodes of "naive empiricism" were usually poorly organized and often interfered with forms of deductive verification that students knew how to use. Finally, Kintsch & Greeno (1985) described a process model of solving arithmetic word problems in which quantitative strategies were triggered by information contained in a "situation model" of the problem. The situation model was constructed during text comprehension and contained a set-based representation of typed quantities and their interrelationships (e.g., part-whole). Follow-on studies (Kintsch, 1986) have shown that the construction of a situation model is important for recall, inference, and learning from text.

Looking over this evidence, we find that studies of mathematical problem solving consistently encounter activities similar to what we call model based reasoning: subjects construct some form of situation model, take inferences within the model to help comprehend and sometimes to solve a quantitative problem, and use the model in a supportive role for assembling or verifying quantitative constraints. Beyond model based reasoning in mathematical problem solving, similar evidence is available across a wide range of cognitive activities. For example, Johnson-Laird (1983) argues for a model driven account of syllogistic reasoning that underlies common sense inference. Given a pair of premises like, *All the artists are beekeepers/All the beekeepers are chemists*, Johnson-Laird's subjects appear to build successively more elaborate models of the situation described by the premises when searching for valid inferences. The validity of each inference, rather than being logically deduced by sound rules of inference, is evaluated with respect to these concrete models of the premises. Errors occur when subjects are unable to build sufficient models of the premises and thus overlook or fail to eliminate various inferences. Relatively concrete forms of reasoning outside traditional (i.e., schooled) formalisms have also been observed for decision making under uncertainty (Tversky & Kahneman, 1974), various forms of statistical reasoning (Nisbett, Fong, Lehman, & Cheng, 1987), and explanations of physical processes (Clement, 1983; McCloskey, 1983).

In general, these studies raise questions about the relationship between what students bring to an educational setting i.e., their "preconceptions" about a subject matter and materials that the curriculum explicitly presents. In the domain of mathematical problem solving, students' "preconceptions" and associated activities are often pushed to the background of legitimate practice and inquiry. At best they are "auxiliary" to quantitative reasoning, while at worst they interfere with preferred problem-solving activities and produce "lost opportunities, unfocused work, and wasted effort" (Schoenfeld, 1985, p. 308). In their stead, the manipulation of symbolic representations of quantity, quite apart from the situations that give rise to these quantities, is held in the foreground. Our findings on model-based reasoning, in concert with other studies reviewed briefly above, suggest that this foreground/background conception of quantitative problem solving may need to be reconsidered.

In our sample of "competent" subjects, a routine problem is one in which the use of familiar algebraic operations will provide a *precise* value for an unknown entity. This is the power of the algebraic formalism: it is perfectly general, sound, and often simple to apply. However, quantitative precision is of little value when the subject is uncertain about the problem's structure. Our characterization of overall episodic activity, the frequency and consequence of conceptual versus manipulative errors during those episodes, and the role of model-based reasoning show that routine activities within the algebraic formalism make up only a portion of competent problem-solving. For many of our subjects, algebra story problems are not routine exercises. Instead, much of their problem-solving activity is devoted to assembling a sensible set of constraints on a desired quantity, an effort that uncovers the problem's structure. When algebraic constraints are unclear, subjects sometimes attempt solutions using model-based reasoning (e.g., Figure 8), a tactic that approximates a *certain* value for an unknown entity. The value is certain when quantitative constraints that determine its derivation are grounded in a representation of problem structure that is

familiar to the subject.

The strategic significance of this activity is consistent with varying explanations. On one hand, *enacting* a set of physical constraints may provide otherwise skilled quantitative problem solvers with an efficient means of estimating quantitative solutions. Under this interpretation, the model based episode shown in Figure 8 may result simply from the subject's preference for repeated additions over a more complicated division. Wilkening (1981) makes a similar argument when interpreting results of a developmental study on the relationship between velocity, time, and distance. In contrast, we argue that episodes of model-based reasoning serve as problem solving strategies in their own right, and are used when more "formal" activities (e.g., algebraic substitution) are unreachable given the current problem representation. Under this interpretation, the subject in Figure 8 undertakes model-based reasoning because her representation of the problem cannot justify a division of the total distance by a combined rate. Enacting motion and time constraints over successive hours of travel makes the quantitative structure of the problem more certain. The results of model-based reasoning support a conceptualization of quantitative constraints in which the total distance can be divided by a *combined rate* to give a *precise account* of the elapsed time. Further constraints are introduced by establishing that the correct quantitative solution falls between the fifth and sixth hours of travel.

Interpreting model-based reasoning as an alignment of certain and precise representations of problem structure leads to deeper questions about a competent understanding of mathematical concepts, in this case related linear functions. One point of view takes mathematical concepts as objects of knowledge in and of themselves, quite apart from their physical embodiment in a situational context. Hence the story in an algebra story problem serves only as a vehicle for carrying a mathematical structure. An alternative point of view takes mathematical concepts as tools for modeling physical situations, in this case related motion or work events as presented in problem texts. The question is how far vehicles

will travel or how long it will take to complete a job, and mathematical concepts serve as sometimes useful tools for answering these questions.

We suspect that these points of view are not incompatible. In fact, the latter view may provide an educational bridge to mathematical concepts as self contained sources of knowledge. That is, a competent mathematical conception of related linear functions is *based on* and *extended through* a physical understanding of the situational context that the "story" of an applied problem presents. An activity like iterative simulation "joins" concrete inclines, allowing the subject to successively construct a systematic relationship between rates and providing an introduction to related linear functions that can be directly supported within a familiar context. Over time, the mathematical concept reflects a history of use as a tool for modeling physical situations. The concept of rate changes as its modeling role is extended over a wider range of situational contexts, perhaps leading to heuristic estimates or algebraic constructions based on "scaling" inclines as invariant relations. The result could eventually resemble a relatively context-free mathematical abstraction. Of course, this account of the acquisition of mathematical concepts is highly speculative and not a focus of our study. However, judging from the problem–solving behavior observed in this study, even ostensibly "competent" mathematical problem solvers continue to base their quantitative efforts within the situational context of presented problems.

Educational implications

We have interpreted the relative prevalence and consequence of conceptual versus manipulative errors as evidence that subjects have difficulty in assembling the quantitative structure of algebra story problems, long after they have mastered the algebraic formalism. Likewise, the prevalence and functional role of model-based reasoning are interpreted as evidence that even mathematically-sophisticated problem solvers explore the situational context of these problems in an attempt to construct or repair a representation that will support a solution. Based on these findings and their interpretation, we examine several implications

for teaching mathematical problem solving.

The primacy of conceptual errors and use of model based reasoning, in some cases to recover from these errors, suggest that instruction based solely within the mathematical formalism may be inadequate for solving non routine problems. Textbook instruction in algebra story problem solving typically addresses this issue by providing some suggestions for "... translating from words to appropriate algebraic forms" (Kolman & Shapiro, 1981, p. 64). These range from direct translation rules taking textual phrases into equations (e.g., rewrite "twice" as 2×) to the construction of tables that organize quantitative entities and their interrelationships around known formulas. The desired result is a set of simultaneous linear equations amenable to algebraic operations. While these suggestions provide a sort of organizational strategy for the student's problem-solving activity, they fall well short of specifying how quantitative relations, particularly those that are only implied by the problem text, can be identified, arranged as entries in a table, or effectively used. Instead, the results of our study point to persistent problem solving difficulties that the traditional algebra curriculum addresses weakly if at all.

How might these components of competent problem solving be taught more effectively? We argue that the situational context of an algebra story problem, and in particular the correspondence between situational relations and quantitative constraints, should be a legitimate object of teaching in the algebra curriculum. This is clearly appreciated in other problem-solving curricula. For example, consider the utility of force diagrams for solving statics problems in physics. Students who ignore or incorrectly construct force diagrams can be expected to manipulate equations or formulas without visible signs of progress. This is quite similar to Paige & Simon's (1966) finding that "auxiliary representations" helped subjects to detect impossible algebra story problems, sometimes before writing any equations at all. Our question, then, is whether there might not be a similar organizing representation for algebra story problem solving? There are some suggestive precedents:

Gould & Finzer (1982) describe an animated computational environment that allows students to make guesses in a one-dimensional world of motion; Greeno (1983) describes an effective instructional technique in which students use an electric train set to help calculate solutions to compound motion problems.

As one possibility among many, we present a representation that draws directly from the analysis of situational structure presented earlier and consider under what circumstances it could provide a useful instructional model for constructive problem solving. As with any model used in teaching, there are problems of registration: the model may cover some aspects of the target domain well but cover other aspects poorly. Our proposal addresses relations and operations possible within a representation of the situational structure of compound algebra story problems, and the correspondence of these aspects to relations and operations possible with a representation of quantitative structure. We expect that in combination with a quantitative model like that proposed by Greeno *et al.* (1986), their joint contribution could prove more effective than either used alone.

[Insert Figure 15 about here.]

Figure 15 shows paired graphical representations of situational and quantitative structure for the *MRT* problem. At the top of the figure, a *dimensional frame* displays orthogonal output (in this case, distance) and time dimensions, with entities arranged along those dimensions by their respective situational relations: times are *adjacent* and distances *congruent*. At the bottom of the figure, a quantitative network (Shalin & Bee, 1985) shows the common distance found by applying motion rates to component times. Each representational device provides a directly accessible illustration for important aspects of competence in this problem-solving domain.

In contrast with translation rules or tabular arrangements, the illustrative medium of dimensional frames provides a spatial abstraction for compound rate problems that promotes a physical justification for essential quantitative constraints. Time segments

add because they are *adjacent* within the vertical dimension, while distance segments are equal because they are *congruent* within the horizontal dimension. As noted in our earlier discussion of quantitative structure, substructures corresponding to these constraints must be constructed before using the quantitative network to find a solution e.g., the additive triad over time extensives that centers the quantitative network in Figure 15. The ability to appropriately select and place these quantitative substructures appears to require a substantial investment in training time (Greeno *et al.*, 1986). We expect that a well designed illustration[9] around the idea of dimensional frames could effectively support the acquisition and use of a quantitative network illustration.

In contrast with a set of algebraic equations, quantitative networks provide a spatial abstraction for variables and equivalence relations that makes the global structure of what would otherwise be a linear encoding more apparent. Rather than writing a set of equations with repeated variable names or constants, a notation that can obscure the role of quantitative entities and make the applicability of certain algebraic operations difficult to recognize, the quantitative network directly captures the notion of shared variables or constants and multiple ways of reaching a particular unknown. The network provides a visually inspectable form of algebraic calculus, essentially *constraint propagation*, that may prove easier for students to learn than more traditional instructional methods (i.e., algebraic operations on linear equations). Thus, the two illustrative media are collaborative in that they provide interdependent representational stages intermediate between a problem text and a correctly manipulated set of algebraic constraints.

Returning to Figure 15, we give a more detailed treatment of this collaborative interdependence. As a compound motion problem, *MRT* involves two events, each contributing entities modeled as segments on output and time dimensions. Across events, segments on each dimension are related in a manner that determines their quantitative composition.

[9]Ohlsson (in press) gives a prescriptive methodology for constructing *interactive illustrations* as well as a particular illustration, called "Rectangle World," for the ratio sense of rational numbers

Adjacent time segments can be composed to yield a single segment whose extent along the vertical dimension corresponds directly to the value of total time traveled, thus implying an additive relation over extensives in the quantitative network. Similarly, *congruent* segments in the distance dimension have an identical extent, implying the same (and same-valued) extensive in the quantitative network. Within each event, the rate provides a comparative mapping between dimensions, modeled as individual inclines in the figure. Placed at the top of the dimensional frame, walking covers 3 miles in one hour, and after transformation to reflect a common output (discussed in a moment), the bus is shown to cover the same 3 miles in $\frac{1}{8}$ hours at the bottom of the frame.

In addition to sanctioning relations among quantitative entities, more direct problem-solving inferences using model-based reasoning are also possible within the dimensional frame. Treated as invariant relations across dimensions, motion inclines can be "scaled" to give heuristic estimates of common distance and composed times, as shown with dashed lines in Figure 15. Alternately, treating rates as concrete associations, inclines could be "joined" together during an iterative simulation of compound motion. In each case, a model-based solution is reached when a common distance is found that precisely requires six hours for round trip traversal. Both forms of model-based solution attempts are consistent with observed protocols. For example, subject m31 uses a form of "scaling" to make heuristic estimates of 24, 12, and 15 miles for a common distance, checking the combined time *required for each estimate against the given six hours*. After the third estimate, she notices that "each mile takes... $\frac{9}{24}$ hours" and later uses this constraint to construct an algebraic expression in a single unknown, "$\frac{9}{24} \times X = 6$." In contrast, subject m18 uses a form of "joining" by choosing 3 miles as a concrete distance segment, determining that the bus takes 7.5 minutes to cover this distance (shown as $\frac{1}{8}$ hours in Figure 15), and then extending these concrete relations in a simulation of successive three-mile return trips. Both subjects alter the form in which motion rates are expressed (i.e., output over time) during their model-based solution attempts, and subject m31 finds a way of combining rates for a "return trip mile." *In each case, activities within model based reasoning episodes observed in written* protocols directly sanction multiplicative relations between rates (intensives) and times (extensives) shown in the quantitative network of Figure 15.

An appropriate combination of these representations could be a helpful artifact for instruction in algebra story problem solving. First, representational choices in the dimensional frame can serve as justifications for more abstract relations or operations in the quantitative network. As argued above, a justification for adding times within the quantitative formalism is that their composed spatial extent is sensible within the situational context of the story. As a more complex example, subject m31's decision to transform and then add motion rates in this problem cleverly restructures the dimensional frame to have single segments on both time and output dimensions -- e.g., $\frac{9}{24}$ hours for each "return trip" mile. The corresponding quantitative network would require only three entities: a time extensive (6 hours, given) results from multiplying the combined rate intensive ($\frac{9}{24}$ hours per mile, inferred) by an unknown extensive for round trip distance. This is a sensible change in representation only because the time segment given in the "goal state" of the problem is presented as a composed whole (i.e., "... he was gone for 6 hours" in the text of problem *MRT*), and round *trip distance segments are congruent. Thus, representational choices* in the dimensional frame provide justification for construction of a simplified quantitative network.

Second, problem-solving activity (e.g., iterative simulation) within the dimensional framework can actually help to recover from prior conceptual errors. For example, consider a subject who first attempts a solution within the algebraic formalism and omits the constraint that distances are the same (i.e., the same variable). Finding two simultaneous linear equations in three variables, this subject reaches an impasse. Choosing model based reasoning for the purpose of comprehension in the next episode, the subject immediately

faces a representational decision in the distance dimension: should positionally distinct or identical spatial segments be chosen? Certainly, the possibility of an incorrect choice remains, but when making this choice in the algebraic formalism of the prior episode, the consequences of an incorrect representational decision were less apparent. Correctly choosing *congruent* distance segments in the dimensional frame could allow this subject to achieve a solution within the model-based reasoning episode, or to return to the algebraic formalism with a more complete representation.

In summary, choosing an apt combination of situational and quantitative models for instructional purposes is a challenging problem. Our suggestion for the dimensional frame as an illustrative mechanism would require further refinement to achieve effective integration with an algebraic illustration, as discussed above. Nonetheless, we feel this approach is interesting in several respects. First, our proposal is consistent with an empirical picture of episodic problem-solving behavior in mathematically sophisticated subjects. Taking these findings as evidence for competent (if not expert) problem solving, we are interested in supporting what problem solvers actually do during their attempts to solve non-routine problems. Our instructional proposal is based on a characterization of these attempts and an analysis of common problem-solving difficulties. Second, although the solution of a particular class of problems may become routine with practice, the ability to construct an algebraic representation will continue to be important for novel problems or problems that have become unfamiliar with the passage of time. Being able to construct a representation in the algebraic formalism, based on the *constraint-generating* inferences we have described as one role for model-based reasoning, may never become entirely routine. Last, combined illustrative media may be of some practical value in delivering instruction on algebra story problem solving, whether provided through computer-based instruction or a traditional algebra curriculum.

ACKNOWLEDGEMENTS

This paper is an expanded version of an earlier report on *the episodic structure of algebra story problem solving*. The research was primarily supported by the Cognitive Science Program of the Office of Naval Research, contract N00014 85 K 0373. Rogers Hall was also supported by a University of California Regents' Dissertation Fellowship; Etienne Wenger was also supported by the Institute for Research on Learning; and Chris Truxaw was also supported by a graduate fellowship from the Office of Naval Research and the American Society for Engineering Education. We wish to thank Jim Greeno, Lauren Resnick, and Karen Wieckert for valuable editorial comments.

REFERENCES

Allen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communications of the Association for Computing Machinery, 26*(11), 832-843.

Allen, J.F. (1984). Towards a general theory of action and time. *Artificial Intelligence, 23*(2), 123-154.

Bobrow, D.G. (1968). Natural language input for a computer problem-solving system. In M. Minsky (Ed.) *Semantic information processing* (pp. 135-215). Cambridge, MA: MIT Press.

Brachman, R.J. (1979). On the epistemological status of semantic networks. In N.V. Findler (Ed.) *Associative networks: Representation and use of knowledge by computers* (pp. 3-50). New York: Academic Press.

Carraher, T.N., Carraher, D.W., & Schliemann, A.D. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education, 18*(2), 83-97.

Carraher, T.N., & Schliemann, A.D. (1987). Culture, arithemetic and mathematical models. Universidade Federal de Pernambuco, Recife, Brazil, manuscript.

Chi, M.T.H., Bassoc, M., Lewis, M.W., Reimann, P., & Glaser, R. (1987). Self-explanations: How students study and use examples in learning to solve problems. Technical Report No. 9, University of Pittsburgh, Learning Research and Development Center.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A.L. Stevens (Eds.) *Mental models* (pp. 325-341). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

de la Rocha, O.L. (1986). Problems of sense and problems of scale: An ethnographic study of arithmetic in everyday life. Ph.D. Dissertation, University of California, Irvine.

de Kleer, J. (1977). Multiple representations of knowledge in a mechanics problem solver. *Proceedings of the International Joint Conference on Artificial Intelligence*, 299-304.

de Kleer, J. (1979). The origin and resolution of ambiguities in causal arguments. *Proceedings of the International Joint Conference on Artificial Intelligence*, 197-203.

Dellarosa, D. (1985). Abstraction of problem type schemata through problem comparison. Technical Report No. 146, Institute of Cognitive Sciences, University of Colorado.

Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Gould, L., & Finzer, W. (1982). A study of TRIP: A computer system for animating time-rate distance problems. *International Journal of Man Machine Studies, 17*, 109-126.

Greeno, J.G. (1983). Conceptual entities. In D. Gentner & A.L. Stevens (Eds.) *Mental models* (pp. 227-252). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Greeno, J.G. (1985). Advancing cognitive science through development of advanced instructional systems. Paper presented at the AERA meetings, Chicago, IL.

Greeno, J.G. (1987). Instructional representations based on research about understanding. In A. Schoenfeld (Ed.) *Cognitive science and mathematics education* (pp. 61-88). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Greeno, J.G., Brown, J.S., Foss, C., Shalin, V., Bee, N.V., Lewis, M.W., & Vitolo, T.M. (1986). Cognitive principles of problem solving and instruction. Technical Report No. 41, Berkeley Cognitive Science Report Series.

Hall, R.P. (1987). When trains become hoses: Justifying problem comparisons in problem solving. Paper presented at the *Third International Conference on Artificial Intelligence and Education*, University of Pittsburgh and Learning Research and Development Center, Pittsburgh, Pennsylvania.

Hall, R.P. (1988). Painting boards and running marathons: Learning to construct an algebraic representation of related linear functions. University of California, Irvine, working paper.

Hinsley, D., Hayes, R., & Simon, H. (1977). From words to equations: Meaning and representation in algebra word problems. In P. Carpenter & M. Just (Eds.) *Cognitive processes in comprehension* (pp. 89–106). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Johnson Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge: Harvard University Press.

Kilpatrick, J. (1967). Analyzing the solution of word problems in mathematics: An exploratory study. (Ph.D. Dissertation, Stanford University.) University Microfilms, 68-6442.

Kintsch, W. (1986). Learning from text. *Cognition and Instruction, 3*(2), 87–108.

Kintsch, W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129.

Kolman, B., & Shapiro, A. (1981). *College algebra and trigonometry.* New York: Academic Press.

Lave, J. (1986). The values of quantification. In J. Law (Ed.) *Power, action and belief* (pp. 88–111). London: Routledge and Kegan.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life* Cambridge: Cambridge University Press.

Lucas, J.F. (1980). An exploratory study on the diagnostic teaching of heuristic problem solving strategies in calculus. In J.G. Harvey & T.A. Romberg (Eds.) *Problem solving studies in mathematics* (pp. 67–91). University of Wisconsin Madison: Wisconsin Research and Development Center for Individualized Schooling, Monograph Series.

Mayer, R.E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instructional Science, 10,* 135–175.

Mayer, R.E., Larkin, J., & Kadane, J. (1984). A cognitive analysis of mathematical problem-solving ability. In R. Sternberg (Ed.) *Advances in the psychology of human intelligence, 2* (pp. 231–273). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

McCloskey, M. (1983). *Naive theories of motion.* In D. Gentner & A.L. Stevens (Eds.) *Mental models* (pp. 325–341). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Newell, A. (1969). Heuristic programming: Ill structured problems. In J.S. Aronofsky (Ed.), *Progress in Operations Research, 3,* 360–414.

Nisbett, R.E., Fong, G.T., Lehman, D.R., & Cheng, P.W. (1987). Teaching reasoning *Science, 30,* 625–631.

Novick, L.R. (1987). Problem similarity, analogical transfer, and individual differences. University of California at Los Angeles, Unpublished manuscript.

Ohlsson, S. (in press). Sense and reference in the design of interactive illustrations for rational numbers. In R. Lawler & M. Yazdani (Eds.), *Artificial intelligence and education.* Norwood, New Jersey: Ablex.

Paige, J.M., & Simon, H.A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.) *Problem solving: Research, method, and theory* (pp. 51-119). New York: John Wiley.

Pirolli, P.L., & Anderson, J.R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology 39*(2), 240-272.

Polya, G. (1945). *How to solve it.* Princeton, New Jersey: Princeton University Press.

Quintero, A.H. (1981). The role of semantic understanding in solving multiplication word problems: Two step R x E problems. In Schwartz, J.L. *The role of semantic understanding in solving multiplication and division word problems.* Final report to the National Institute of Education, NIE-G-80-0144.

Quintero, A.H., & Schwartz, J.L. (1981). The development of the concept of ratio in children. In Schwartz, J.L. *The role of semantic understanding in solving multiplication and division word problems.* Final report to the National Institute of Education, NIE-G-80-0144.

Reed, S.K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(1), 106-125.

Reed, S.K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology; Learning, Memory, and Cognition, 13*(1), 124-139.

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburg (Ed.) *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.

Schoenfeld, A.H. (1985). *Mathematical problem solving.* New York: Academic Press, Inc.

Schwartz, J.L. (1982). The semantic calculator. *Classroom Computer News 2*, 22-24.

Shalin, V., & Bee, N.V. (1985). Analysis of the semantic structure of a domain of word problems. Technical Report No. APS 20, University of Pittsburgh Learning Research and Development Center.

Silver, E.A. (1979). Student perceptions of relatedness among mathematical verbal problems. *Journal for Research in Mathematics Education, 10*, 195-210.

Silver, E.A. (1981). Recall of mathematical problem information: solving related problems. *Journal for research in mathematics education, 12*(1), 54-64.

Simon, H.A. (1973). The structure of ill structured problems. *Artificial Intelligence, 4*, 181-201.

Singley, M.K. (1986). Developing models of skill acquisition in the context of intelligent tutoring systems. Ph.D. Thesis, Carnegie Mellon University.

Thompson, P.W. (1988). Artificial intelligence, advanced technology, and learning and teaching algebra. To appear in C. Kieran & S. Wagner (Eds.), *Research issues in the learning and teaching of algebra.* Reston, VA: National Council of Teachers of Mathematics.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology, 13*, 231-247.

Table 1: Representative algebra story problems.

Motion: Opposite direction (*MOD*).

Two trains leave the same station at the same time. They travel in opposite directions. One train travels 60 km/h and the other 100 km/h. In how many hours will they be 880 km apart?

Motion: Round trip (*MRT*).

George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

Work: Together absolute (*WT*).

Mary can do a job in 5 hours and Jane can do the job in 4 hours. If they work together, how long will it take to do the job?

Work: Competitive (*WC*).

Randy can fill a box with stamped envelopes in 5 minutes. His boss, Jo, can check a box of stamped envelopes in 2 minutes. Randy works filling boxes. When he is done, Jo starts checking his work. How many boxes were filled and checked if the entire project took 56 minutes?

Table 2: Categories for interpreting the purpose, content, errors, and relative status of problem-solving episodes.

- Strategic purpose
  - Comprehension
  - Solution attempt
  - Verification
- Tactical content
  - Annotation
    - Problem elements
    - Retrieval of formulas
    - Diagram
  - Algebra
  - Model-based reasoning
    - Simulation
    - *Heuristic*
  - Ratio
    - Whole/part
    - Part/whole, part/part
    - *Proportion*
    - Scaling
  - Unit
  - Procedure
- Conceptual content
- Errors
  - Conceptual errors
    - *Errors of commission*
    - Errors of omission
  - Manipulation errors
    - Algebraic errors
    - Variable errors
    - Arithmetic errors
- Status of episode within solution attempt
  - Consistency
    - *Before*
    - During
    - After
  - Transition
    - Subgoal
    - Wrong
    - Impasse
    - Lost
    - Final solution
    - Found solution wrong

Table 3: Percentage of subjects with a scored category during their solution attempts.

Problem	MOD	MRT	WT	WC
Tactical content				
Algebra	82.4	85.9	71.8	63.5
Model	30.6	22.4	35.3	47.1
Ratio	17.6	14.1	15.3	42.4
Procedure	0.0	1.2	21.2	0.0
Units	3.5	1.2	1.2	1.2
Notations	7.1	15.3	21.2	29.4
Diagram	69.4	36.5	8.2	9.4
Strategic purpose				
Comprehension	84.7	64.7	57.6	60.0
Solution attempt	100.0	100.0	100.0	100.0
Verification	28.2	20.0	7.1	20.0
Episode transitions				
Solution	97.6	75.3	85.9	97.6
Impasse	9.4	10.6	7.1	4.7
Lost	4.7	21.2	15.3	3.5
Wrong	16.5	38.8	25.9	16.5
Errors				
Omission	7.1	21.2	23.5	11.8
Commission	17.6	49.4	42.4	14.1
Arithmetic	9.4	4.7	3.5	2.4
Algebra	5.9	8.2	8.2	0.0
Variable	1.2	5.9	14.1	2.4

Table 4: Final episodes: percentage correct by subject groupings.

Problem	MOD		MRT		WT		WC	
Group contrast[a]	M < W		M > W		M > W		M < W	
Correct	89.1	92.3	47.8	56.4	58.7	64.1	93.5	89.7
Incorrect	6.5	7.7	19.6	15.4	28.3	20.5	6.5	5.1
No-solution	4.3	0.0	32.6	28.2	13.0	15.4	0.0	5.1

[a] M sees *MOD, WT, WC, MRT*; W sees *WT, MOD, MRT, WC*.

Table 5: Final episodes: tactical content and errors by correctness.

Problem	MOD			MRT			WT			WC		
Outcome[a]	C	I	N	C	I	N	C	I	N	C	I	N
n	77	6	2	44	15	26	52	21	12	78	5	2
Tactical content												
Algebra	58	6	0	36	8	20	43	5	7	44	2	1
Model	3	0	0	4	2	6	2	1	2	12	1	0
Ratio	13	0	2	4	3	0	5	3	2	22	1	1
Procedure	0	0	0	0	0	0	1	11	1	0	0	0
Units	2	0	0	0	0	0	0	0	0	0	0	0
Not scored	1	0	0	0	2	0	1	1	0	0	1	0
Errors												
Conceptual	1	6	0	0	14	16	1	27	10	1	4	0
Manipulative	7	2	0	1	5	2	4	7	1	2	1	0

[a]C = correct; I = incorrect; N = no solution.

Table 6: Errors and transitional status of a previous episode compared with the purpose of a model-based reasoning episode.

Problem	MOD			MRT			WT			WC		
n	26			19			30			40		
Purpose[a]	C	S	V	C	S	V	C	S	V	C	S	V
No preceding episode	7	1	0	1	4	0	17	4	0	10	2	0
No errors in preceding episode												
On track	3	9	0	0	6	0	2	2	1	10	11	2
Abandon	1	0	0	1	1	0	0	2	0	0	0	0
Errors in preceding episode												
On track	1	0	0	0	0	0	0	0	0	1	1	0
Abandon	2	2	0	1	5	0	0	2	0	1	2	0

[a]C = comprehension; S = solution attempt; V = verification.

Table 7: Errors before and during model-based reasoning.

Problem	MOD		MRT		WT		WC	
n	26		19		30		40	
Model episode	Errors	None	Errors	None	Errors	None	Errors	None
Previous episode								
No episode	0	8	2	3	1	20	0	12
Errors	1	4	2	4	0	2	1	4
No errors	0	13	1	7	0	7	0	23

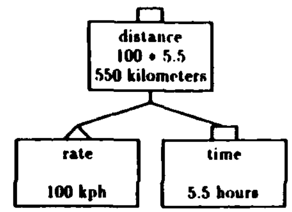

Figure 1: A multiplicative relation involving two extensives and a single intensive.

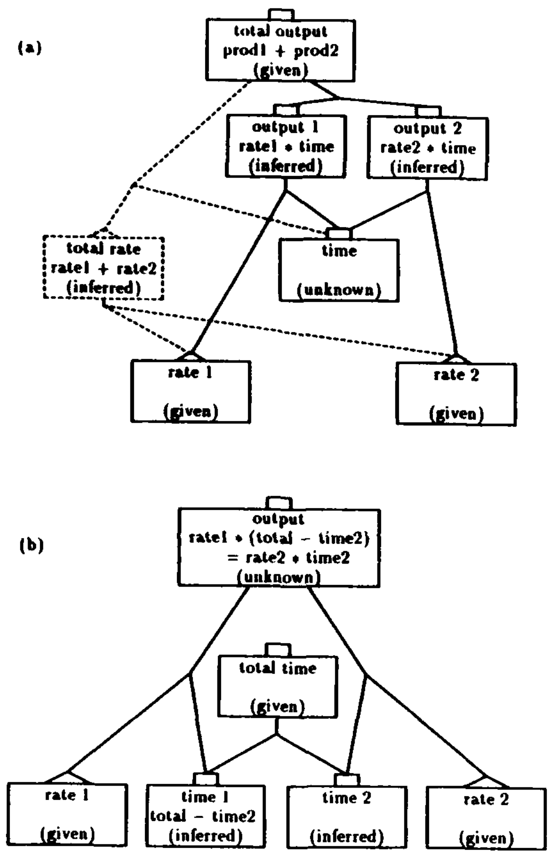

Figure 2: The quantitative structure of two problem classes: (a) contains problems *MOD*, *WT* while (b) contains *MRT*, *WC*.

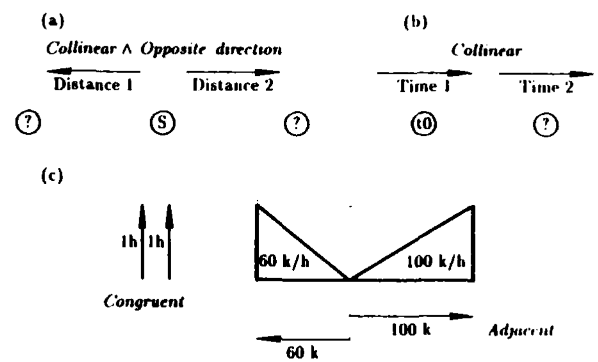

Figure 3: A situational context for motion in opposite directions: (a) and (b) show places and segments for output and time, while (c) shows inclines for rates when these dimensions are arranged orthogonally.

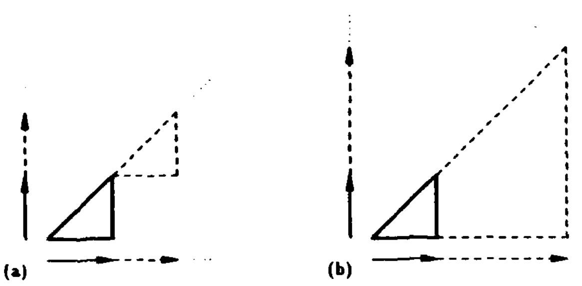

Figure 4: Operations based on different interpretations of two-dimensional inclines: (a) shows a concrete situation successively "joined" to give an iterative simulation of states within the problem model; (b) shows an invariant relation "scaled" to give a heuristic estimate of a final state in the model.

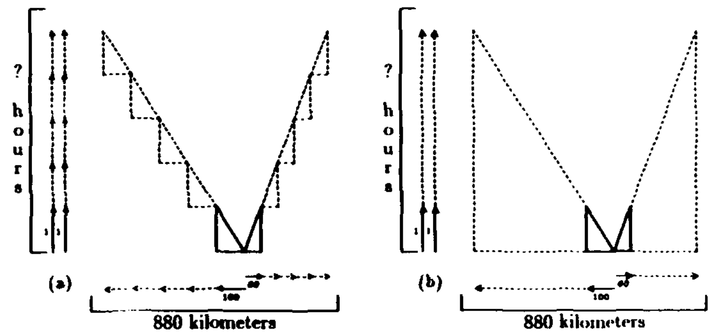

Figure 5: Solution attempts using model-based reasoning on problem MOD: (a) "joins" successive concrete inclines in an iterative simulation; (b) "scales" inclines as an invariant multiplicative relation in a heuristic estimation.

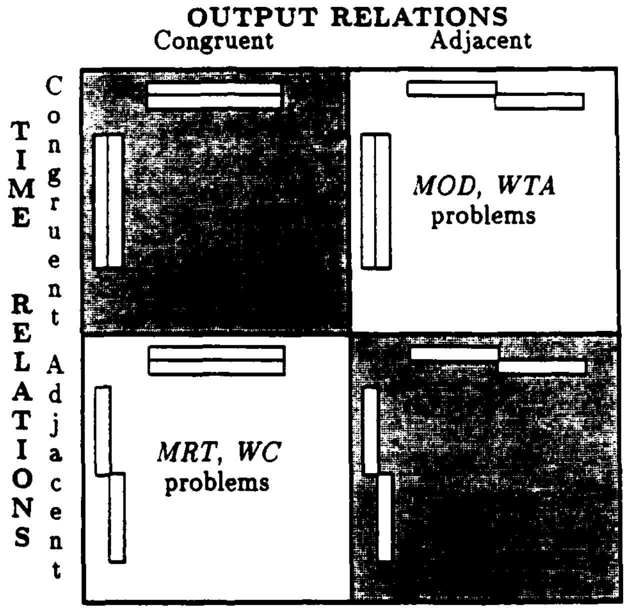

Figure 6: A matrix of situational contexts for pairs of isomorphic motion and work problems.

Mary can do a job in 5 hours and Jane can do the job in 4 hours. If they work together, how long will it take to do the job?

Mary does 1/5 job in 1 hr 1

Jane " 1/4 job in 1 hr

$\frac{1}{5}x + \frac{1}{4}x = 1$

$x\left(\frac{1}{5} + \frac{1}{4}\right) = 1$

$x\left(\frac{4}{20} + \frac{5}{20}\right) = 1$ 2

$x\left(\frac{9}{20}\right) = 1$

$\boxed{x = 20/9}$

DOUBLE CHECK.

$\frac{1}{5}\left(\frac{20}{9}\right) + \left(\frac{20}{9}\right)\frac{1}{4} = 1$

$\frac{4}{9} + \frac{5}{9} = 1$ 3

okay.

Figure 7: Protocol of subject m20 on the WT problem.

Two trains leave the same station at the same time. They travel in opposite directions. One train travels 60 km/h and the other 100 km/h. In how many hours will they be 880 km apart?

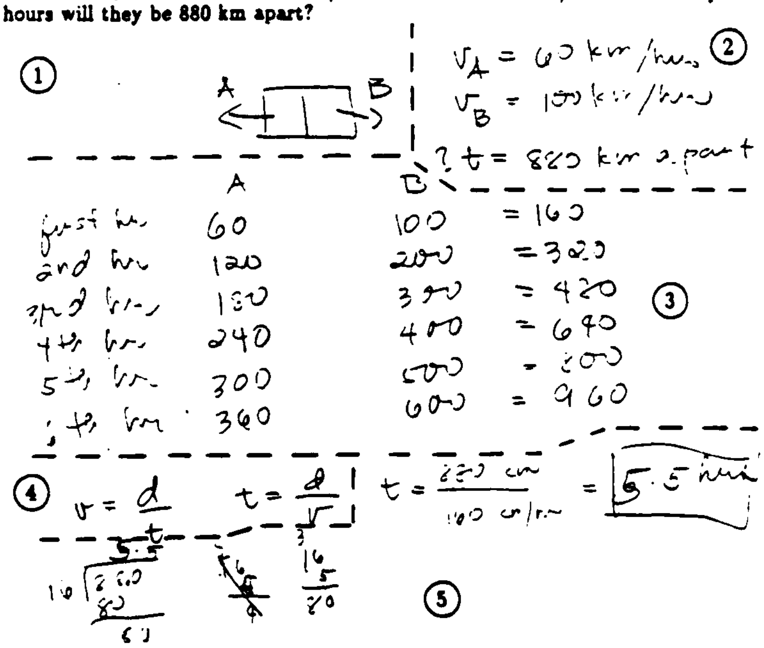

Figure 8: Protocol of subject w06 on the MOD problem.

George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

bus distance = (24 miles/hr)(x hours)

walking distance = (3 miles/hr)(6 − x hours)

bus distance = walking distance

(24 miles/hr)(x hours) = (3 miles/hr)(6 − x hours)

$24x = 18 - 3x$

1 $27x = 18$

$x = \frac{18}{27} = \frac{2}{3}$ hours

bus distance = (24 miles/hr)($\frac{2}{3}$ hours)

bus distance = 16 miles = walking distance

$\boxed{\begin{array}{l}\text{One way} = 16 \text{ miles} \\ \text{Round Trip} = 32 \text{ miles}\end{array}}$

Figure 9: Protocol of subject m39 on the MRT problem.

Two trains leave the same station at the same time. They travel in opposite directions. One train travels 60 km/h and the other 100 km/h. In how many hours will they be 880 km apart?

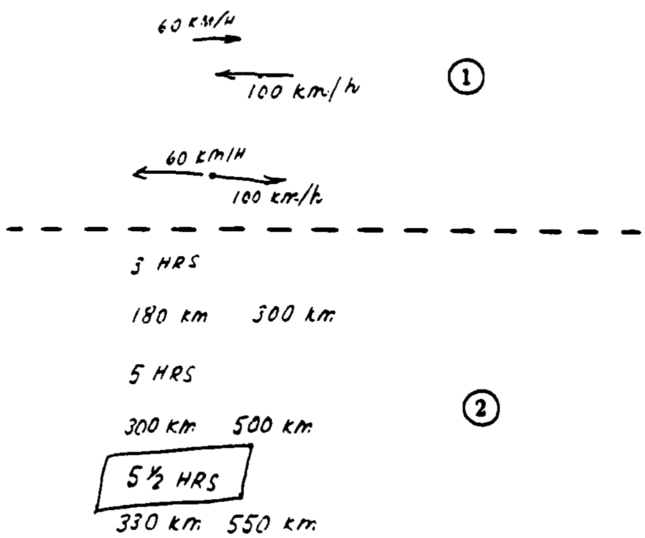

Figure 10: Protocol of subject m03 on the MOD problem.

Mary can do a job in 5 hours and Jane can do the job in 4 hours. If they work together, how long will it take to do the job?

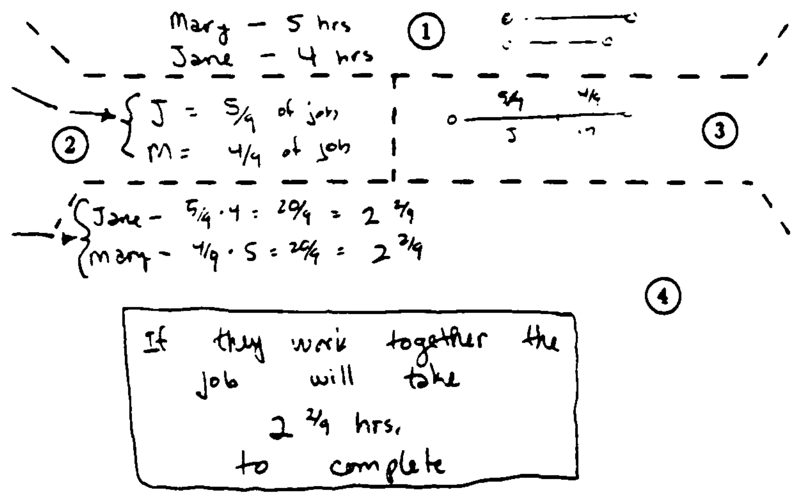

Figure 11: Protocol of subject m32 on the WT problem.

George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

~~$\frac{24\,m}{hr} \cdot x\,hr + 3$~~

1 $\frac{24\,m}{hr} + \frac{3m}{hr}$

$\frac{24\ m/hr}{3\ m/hr} = 8$ 2

Bus travels 8x faster than George

So, If ~~Bus~~ Bus travels 24 miles for one hour, George travels back 24 miles for 8 hours resulting 9 hours total. 3

But we want 6 hours which is $\frac{2}{3} \times 9$.

$24\,m \cdot \frac{2}{3} = 16\,m$ 4

$\boxed{\text{16 miles}}$

Figure 12: Protocol of subject w17 on the MRT problem.

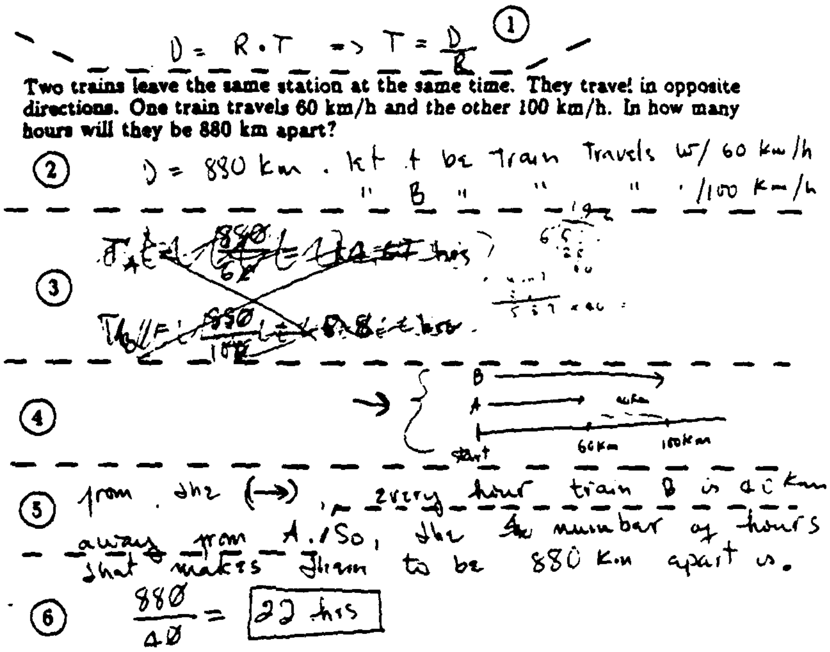

Figure 13: Protocol of subject m19 on the MOD problem.

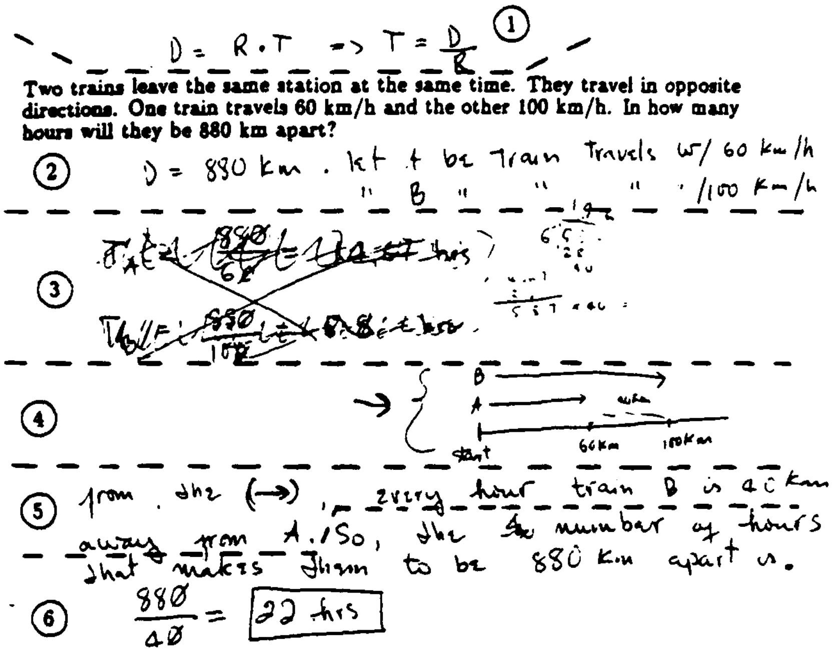

Figure 13: Protocol of subject m19 on the MOD problem.

George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

~~$\frac{24\ m}{hr} \cdot x\ hr + 3$~~

1 $\frac{24\ m}{hr} + \frac{3\ m}{hr}$

2 $\frac{24\ m/hr}{3\ m/hr} = 8$

Bus travels 8x faster than George

So, If ~~Bus~~ Bus travels 24 miles for one hour,
George travels back 24 miles for 8 hours
resulting 9 hours total. 3

But we want 6 hours which is $\frac{2}{3} \times 9$.

$\frac{24\ m}{\cancel{\text{[illegible]}}} \cdot \frac{2}{3} = \frac{16\ m}{\cancel{\text{[illegible]}}}$ 4

$\boxed{\text{16 miles}}$

Figure 12: Protocol of subject w17 on the MRT problem.